A Quest for novel Signaling Molecules in Pleurobrachia Bachei

David Orion Girardo[1] , Mathew Citarella[2], Andrea Kohn[2], Billie J. Swalla[4,5], Leonid Moroz[2,3]

Friday Harbor Labs Marine Genomics Research Apprenticeship

Spring 2012

[1]Departments of Mathematical Sciences, Computer Science, Biology Worcester Polytechnic Institute, Worcester, Massachusetts, USA

[2]The Whitney Laboratory for Marine Bioscience, University of Florida, 9505 Ocean Shore Blvd., St. Augustine, Florida 32080, USA

[3]Department of Neuroscience, University of Florida, Gainesville, Florida 32611, USA

[4]Department of Biology, University of Washington, Seattle, Washington 98195

Contact information:

David Orion Girardo

Departments of Mathematical Sciences, Computer Science, Biology

100 Institute Rd.

Worcester, MA

dogirardo@wpi.edu

Keywords: Pleurobrachia bachei, bioinformatics, neuroscience, neuropeptide

Abstract

The evolution of the nervous system is poorly understood, with broad implications in basic neuroscience and regenerative medicine. The unintuitive nervous system of the basal Ctenophores provides an insight into neuronal evolution. We developed a Zero-Click computational pipeline for secretory peptide prediction from raw genomic, transcriptomic, and proteomic sources (Girardo, 2012), in search of novel signaling pathways. This pipeline was applied to gene models produced from our Pleurobrachia bachei genome, initiated in July 2009 (Moroz, 2012), to obtain a list of 38 putative secretory peptides (Citarella, 2012). 10 of these peptides with no direct homology were selected for cloning and in situ hybridization. Many predicted products localize to neuronal structures, indicating a more complex neuropeptide signaling network than previously described in Ctenophore.

Background

Signaling

Porifera, Placazoa, and Ctenophora form the so called "basal" phyla. Among them, Porifera and Placazoa have no distinguished nervous system (Srivastava, 2008, 2010), leaving the comb jellies ctenophore as the most basal extant metazoan with a true nervous system. Genomic analysis of the ctenophore, *Pleurobrachia bachei*, reveals a lack of many expected classical neurotransmitters such as serotonin, present in their sister phylum, cnideria (Hajj-Ali I, 1997). Remarkably, the only classical neurotransmitter detected in *Pleurobrachia* is glutamate (Citarella, 2011). Further, molecular analysis reveals a wide diversity of glutamate receptor patterns (Swore, 2012), suggesting glutamate as potentially the most ancient neurotransmitter. However, glutamate alone appears insufficient for the complexity of sensory and muscular behaviors exhibited by ctenophores. We hypothesize that ctenophore glutamate signaling is modulated by a rich repertoire of novel neuropeptides, developed through parallel evolution (Moroz, 2009)

Neuropeptides

Classically, neuropeptides are short (typically under 40 Amino Acids) secreted signaling molecules, involved in a wide variety of neuromodulatory effects. There is significantly higher diversity in neuropeptides than in classical neurotransmitters, with over 100 known human neurotransmitters (NCBI, as of 2012). Further, since the only direct action of neuropeptides is to bind to receptors, the biological effect of a neuropeptide is entirely determined by the properties of its receptor. Most neuropeptides act via a G-protein coupled or metabotropic receptors, but the biological function of most neuropeptides is poorly understood or completely unknown.

Machine learning Heuristics

Machine learning is a general-purpose tool of particular use in the analysis of large datasets. Many machine learning techniques are nothing more than sophisticated curve fitting algorithms, which must be 'trained' on known data. Two methods commonly employed in genomic analysis are Hidden Markov Models (HMM) and Artificial Neural Networks (ANN), which are described with great detail by (Stamp, 2004) and (Hopfield,1982) respectively. We employ HMM and ANN extensively in gene prediction, protein family annotation, and protein function prediction. Being *learning* algorithms, they require a substantial set of training data for optimal results. Typically, in the case of protein function and gene prediction, positive and negative examples from the same family are used. In the case of the previously unsequenced phylum Ctenophora, *de novo* prediction necessitates careful modification of model parameters.

Materials and Methods

Prediction

A wide range of prediction methods and features were considered as components to the secreted peptide prediction pipeline. The features considered include an N-terminal signal peptide, a lack of transmembrane domains, cysteine counts, cleavage number, low complexity through repetition and length.

Almost all pre-secreted products are localized to the endoplasmic reticulum via the presence of an N-terminal signal peptide consisting of 5-20 hydrophobic residues. Thus, signal peptide prediction formed the foundation of our prediction pipeline. For signal peptide prediction, we considered and evaluated the tools SignalP (Jannick Dyrløv Bendtsen 2004), TargetP (Olof Emanuelsson 2000),and Protein Prowler(Hawkins 2005). SignalP utilizes separate Hidden Markov Model and Neural Network components, trained on a variety of organisms, for the sole purpose of secretory signal peptide prediction. TargetP utilizes the Neural Network component of SignalP and separate Neural Networks to discriminate between 'secreted', 'mitochondrial', and 'other' localization signals. Protein Prowler improves upon TargetP using a variety of additional heuristic methods. Protein Prowler poses a severe limitation for automated analysis as a web server-only application. Further, Protein Prowler's advanced optimizations primarily focus on plant models and in preliminary comparisons on the *Pleurobrachia* gene models; Protein Prowler was in 99% agreement with TargetP. For these reasons, only SignalP and TargetP were included in the initial pipeline.

Complicating the process is the possibility of transmembrane domains; products containing one or more transmembrane domain that are directed to the ER are not secreted, and instead are embedded in the membrane. TMHMM (Krogh A 2001) is employed to predict transmembrane topology. Unfortunately, transmembrane domains, like signal peptides, are characterized by a region of hydrophobic residues of similar length, often causing false-positive signal peptide or transmembrane results from trained prediction methods such as TMHMM. Phobius (Lukas Käll 2004) uses conflicting HMM, trained separately for transmembrane domains and signal peptides, to call a consensus.

A significant post-translational modification of neuropeptides is release of the active peptide through cleavage. Often, prohormones contain multiple repeated active sites, flanked by cleavage sites. These cleavage sites are typically marked by basic or dibasic residues. Neuropred (Bruce R. Southey 2006) was used with default settings under 'known motif' to predict cleavage sites only. The results were filtered for certainties greater than 70%, and counted to assign each peptide a cleavage site number. Unfortunately, while some common neuropeptides such as FMRFamide have many internal repeats, others such as Neuropeptide Y have none, so low complexity repeats cannot consistently predict signaling molecules.

Propeptides are typically under 200 amino acids in length, and thus length was considered as a prediction feature. However, analysis revealed that short protein length correlates strongly with other prediction features, making length analysis unnecessary.

Cystine production carries a significant biological cost, in both production and hazardous degradation, leading to strict conservation in biologically important molecules.

For this reason, cysteine counts were considered as an additional factor for prediction. However, initial analysis of other filtering methods yields no significant correlation, and, airing on the conservative side, it is omitted in the final pipeline.

While these prediction tools have been tested in a wide range of model organisms, plants, and bacteria, significant validation has not been considered for the basal groups such as ctenophore, and the transferability of these training sets is unclear. For this reason, we have remained cautious in choosing cutoff values, not desiring to falsely cull results due to over fitting. As more experimental data is gained, we continue to re-evaluate our prediction criteria.

The release of SignalP version 4 prompted an upgrade. Since SignalP 4.0 uses a new prediction method for suspected transmembrane proteins, the probability scores do not correspond directly to SignalP 3.0. Thus, the same valley-hill matching produced cutoff points were recalculated to be 0.7, 0.82, 0.91 for least, medium, and most specific sets respectively [Figure4].

Improvements in local database features and the release of SignalP version 4 prompted us to revisit our prediction methods. Review of homology data revealed a likely secreted frizzled-like protein that was excluded from all selection due to a TargetP certainty cutoff of 2. In later predictions, we maintain TargetP statistics for records, but do not use it as a selection criterion. In the future, we may reimplement TargetP with more sophisticated criterion using experimental data.

Informatics analysis

The difference in final predictions between simply removing TargetP criteria, and setting the certainty cutoff to 2 is minimal, due to overlap with SignalP. Additionally, initial tests revealed that changing the TargetP certainty cutoff from 1 to 2 nearly doubled the positive prediction count uniformly across other variables. Thus, it was kept at 1 for all selection sets.

Since internal repeats are unpredictable, little can be said with certainty about the number of cleavage sites. However, it can be said for certain that a neuropeptide must be cleaved once to remove the signal peptide. There is also expected to be at least one other cleave to produce the active peptide.

Phobius compares the probability scores of transmembrane predictions and signal peptide predictions directly, and so does not produce its own probability score. Thus, adherence to 'signal peptide with no transmembrane domain' is strictly enforced for all predictions.

We observed signalP cutoffs to have the greatest effect on prediction numbers. A score of 50% for both the NN and HMM is the default requirement by signalP and TargetP. However, accepting the 50% score yields over 400 results, many more than truly exist. We examined result density over cutoff percentages [Figure2, Figure3] to estimate meaningful cutoff values. In order to achieve an open but tractable initial prediction set, we developed three non-overlapping sets of results using the above reasoning [Table1].

By mapping the predicted genes to our *Pleurobrachia* genome, we obtained intron organization information [Supplement 2]. Using NCBI's public database, we analyzed

the most stringent group via homology search and manually reviewed basic sites and internal repeat motifs [Supplement 3]. Following review, ten candidates were selected and all were successfully cloned [Supplement 1].

Datasets

Secretory predictions were manually reviewed via homology comparison to public databases NCBI, the Joint Genome Institute, and the Broad Institute 'Origins of Multicellularity' project. They were also compared to the *Aplysia californica* genome and transcriptome (Moroz et al, Unpublished) and *Sepia* transcriptome, which together with the *Pleurobrachia bachei* genome (Moroz et al, Unpublished) and transcriptomes are hosted on the Moroz Lab's internal Neurobase server (Citarella, 2011).

Animals

Pleurobrachia bachei were collected from the Friday Harbor Labs docks for *in situ* hybridization. The animals were housed in fresh flowing sea water at approximately 10^{o}C. For *in* situs, animals are fixed in 4% paraformaldehyde in Filtered Sea Water (FSW) overnight at 4^{o}C. After washing 3x 10 minutes in PBST (phosphate buffered saline + 0.1% tween), they are stored in MeOH at -20C up to a week.

Expression analysis

Digoxigenin probes were prepared for each cloned putative secretory peptide. *In situ* hybridization was used to visualize putative peptide expression patterns in adult *Pleurobrachia* [Appendix 1] to gain insight into their biological activity. Duplicate *in situ* procedures were performed on the 10 selected peptides, referred to internally as NP1,

NP2 (GVEDin), NP5, NP7, NP9, NP27, NP33, NP36, NP45 and NP64. To determine the post-translational modifications and secretion status, the predicted sequences were sent for proteomic analysis via Matrix Assisted Laser Desorption Ionization Mass Spectrometry (MALDI MS) (Rohner, 2005), using a protocol designed to select only secreted products.

Results

Informatics analysis

Putative secreted peptides show low homology to related phyla, with only 9 of the 38 being conserved across the basal phyla [Figure 13]. Additionally, they have few introns when mapped to the genome [Figure 12], typical of short, fast evolving genes. By comparing transcriptome data, it is seen that most predicted products are differentially expressed throughout different tissues [Figure 11].

Experimental analysis

Many of the putative peptides show specific staining and unique expression patterns. GVEDin is seen between the combs, the polar field, ciliated furrows, and apical sensory organ [Figure 5]. NP33 appears in a unique single-celled pattern in the mouth [Figure 10]. NP45 and NP64 each are expressed in different single-celled patterns in the polar field [Figure 6]. NP9 generates a dense spotted motif on the tentacle surface [Figure 8]. NP5 is expressed in the nerve cord of the tentacle [Figure 7]. NP27 is expressed diffusely, deep within the tentacle and in the combs. NP36 is similarly found diffusely in the tentacles, while NP1 shows no differential staining. Preliminary MS results indicate the presence of secreted GVEDin.

Discussion

The expression patterns displayed by NP2, NP5, NP45, and NP64 are highly indicative of neural specification described in the literature (Jager, 2011), though the exact group of neurons is poorly described and will require further investigation. Particularly striking are the two parallel nerve cords [Figure 9] associated with NP5. The NP33-associated cells clustering around the mouth are previously undescribed, though their tapering projections are highly indicative of an axonal structure. Further, the evolutionary profile of short, highly divergent, neuron-specific molecules fits the nature of neuropeptides well indicating a likely role in interneuron signaling. However, this profile also fits signaling *receptors* as well, but since these genes were predicted with high certainty by multiple techniques to be secreted, this case is far less likely.

The cell-type localization of NP1, NP36 and NP27 remains unclear. Of note is the lack of detection in the mesogleal nerve-net previously described in ctenophores (Jager, 2011). It was also seen that in NP9, staining halted where the tentacle met the tentacle sheath, which differed depending on how far it was extended when fixed. This suggests that the detection buffer cannot penetrate the mesoglea in sufficient quantity to elicit a response, which may also explain the lack of mesogleal neurons. In future *in situ* experiments, an incision should be made in the mesoglea to allow the detection buffer to permeate.

NP9's expression motif is reminiscent of the previously described colloblast cells [Figure 9]. Given the high expression levels of NP9, it is likely a secreted component of colloblast adhesion used to ensnare prey. While outside the scope of signaling pathways,

identification of colloblasts' adhesion mechanism is an important step in the characterization and annotation of the ctenophore genome and should be pursued further.

While the validation of secreted NP2 via mass spectrometry is promising, more proteomic analysis must be performed to determine the form and function of these peptides. Further, it would be a worthwhile endeavor to attempt a behavioral study by introducing synthesized peptides to elicit a behavioral response. Peptide synthesis is made feasible by the diminutive size of the active peptides; for example, NP2's GVED repeat. However, the complex post-translational modifications must be determined by proteomics first, as existing prediction methods seem incapable of accurate cleavage site localization in such a species so divergent from our training sets.

The wide range of neuronal structures seen, with different neuron types inhabiting the same tissue in the case of NP2, NP45, and NP64, suggests a greater neuronal complexity in ctenophores than previously believed. Indeed, the lack of neuropeptide homologs in other species betrays an extensive history of parallel evolution.

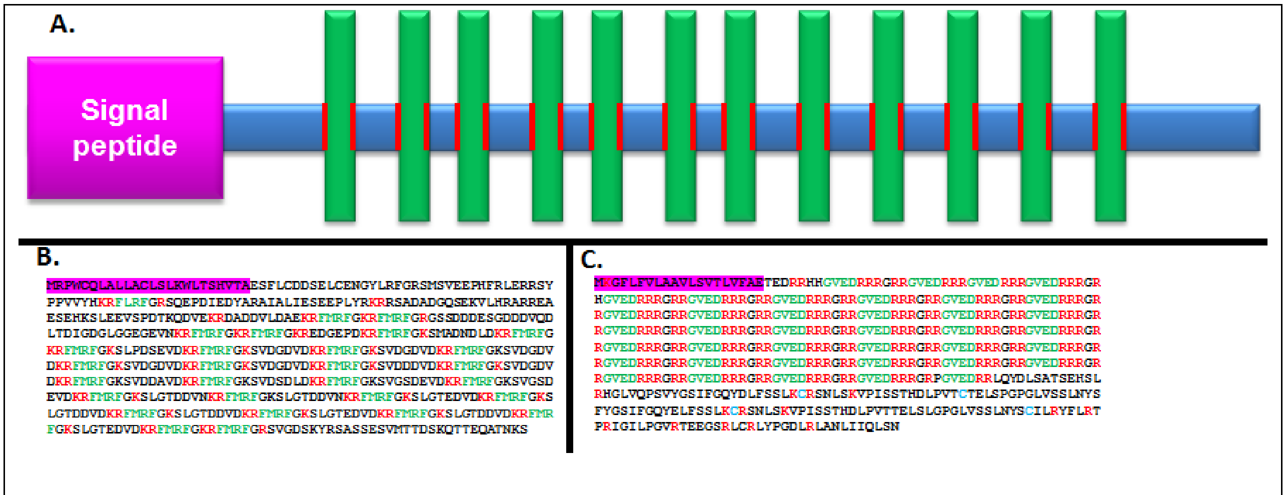

Figure 1. A. The anatomy of neuropeptides. They are characterized by an N-terminal signal peptide, an internally repeated region flanked by basic or dibasic sites. Likewise, they possess no transmembrane domain **B.** A classic example of a neuropeptide, FMRFamide **C.** Predicted *Pleurobrachia* neuropeptide 'GVEDin'

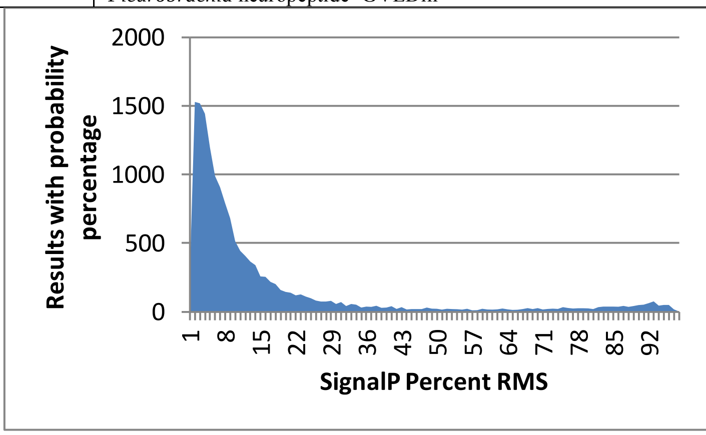

Figure2. SignalP NN and HMM probabilities are combined into a single density plot for Root Mean Square value via RMS = $\sqrt{\frac{1}{2} * (NN^2 + HMM^2)}$. The majority of *Pleurobrachia* proteins have low signal peptide probability. With little deviation towards the center percentages. The area under the curve is the total number of sequences (multiplied by 100)

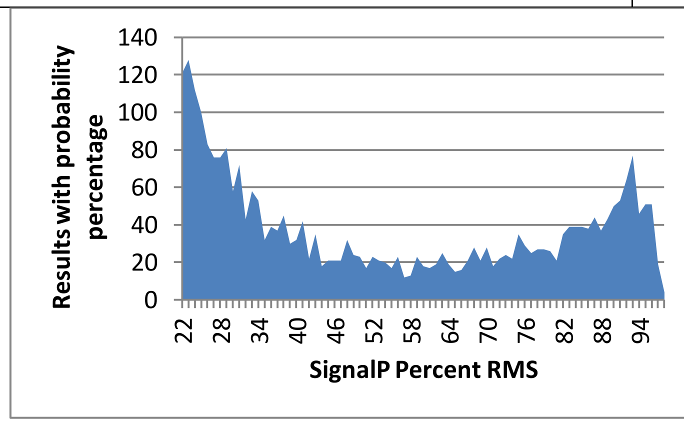

Figure3. Focusing on the tail, there is a second peak beginning after 50% indicating a biologically significant attractor pattern in this region. The curve gains greatest slope just after 90%

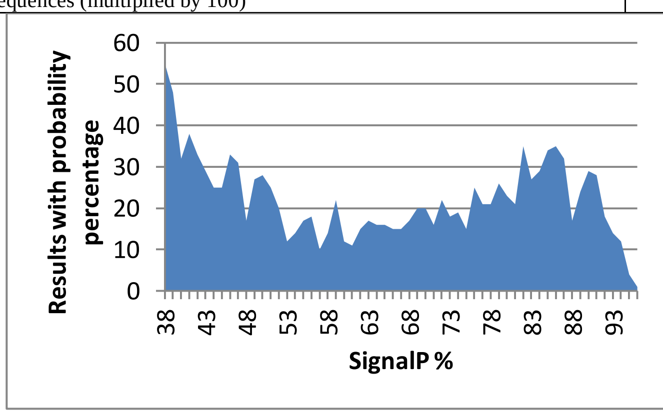

Figure4. SignalP 4.0 uses a single unified prediction percentage. The graph shows a similar motif to 3.0, though shifted slightly and less pronounced.

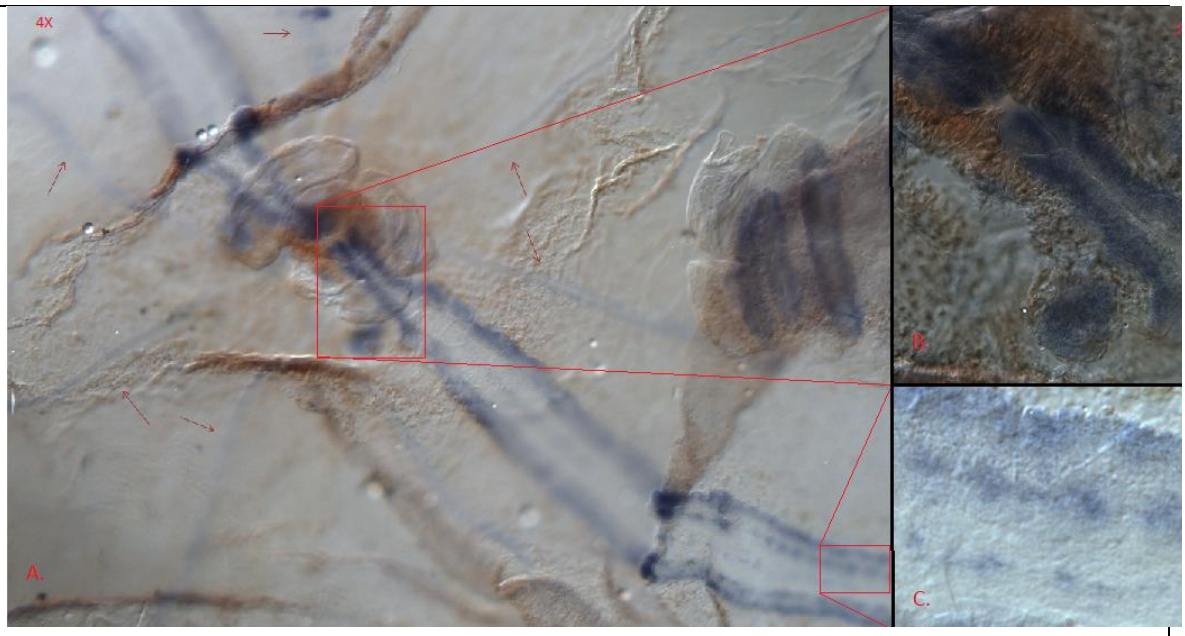

Figure 5. A. GVEDin (NP2) in the aboral end. Expression is evident in the ciliated furrows (Arrows) leading to the combs **B.** Along the apical organ leading up to the statolith. **C.** A distinct, single-celled banding pattern along the polar field.

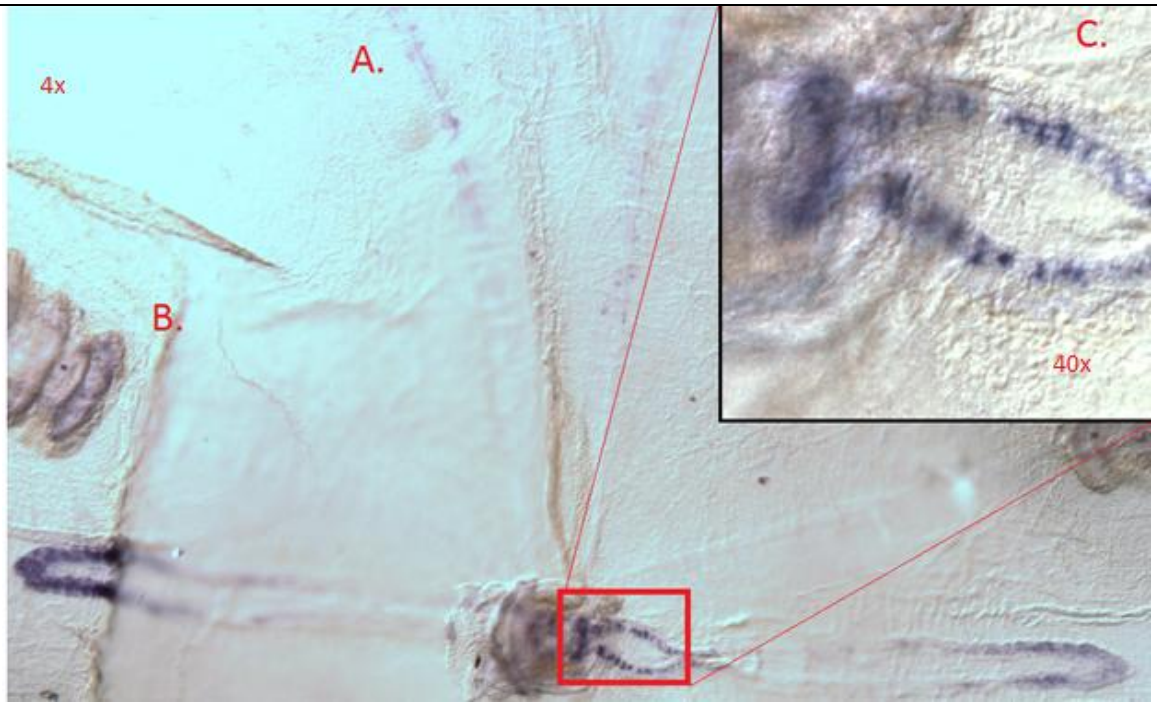

Figure 6. NP64 localization from the aboral end. Expression is seen **A.** in a sparse, single-celled network along the ciliated furrows **B.** Single celled pattern at the base of the combs. **C.** Single celled rows along the outer edge of the polar field, leading into the apical organ.

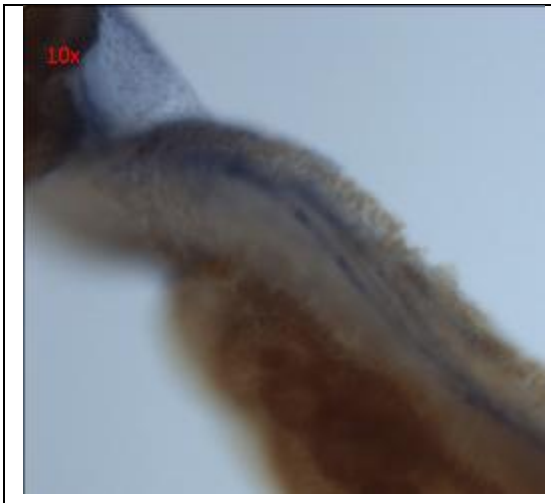

Figure 7. NP5 seen in parallel nerve cords deep in the tentacle

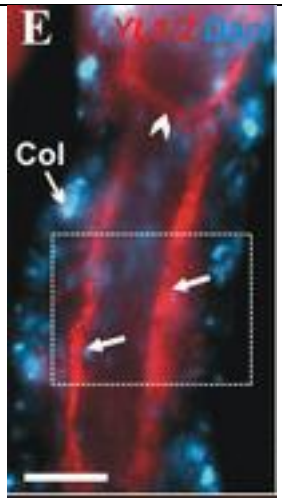

Figure 9. Dapi and YL1/2 immunostaining of the tentacle (Jager, 2010). Displaying colloblasts and two parallel nerve cords through the tentacle center.

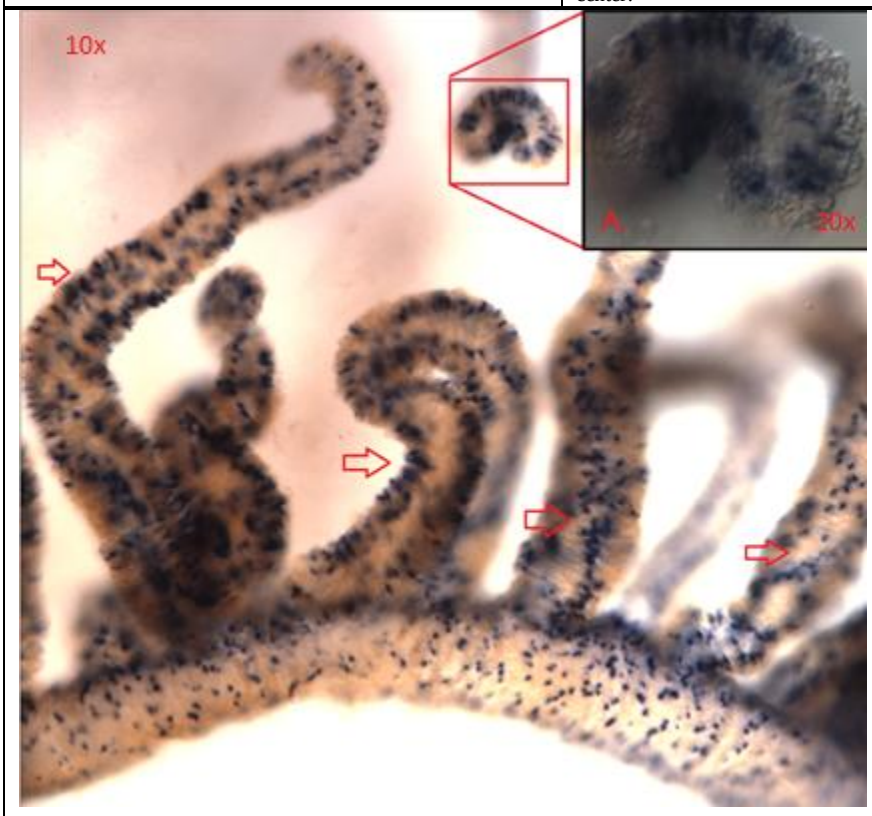

Figure 8. NP9 localization in the tentacle. Expression is seen in individual surface cells. The cells cluster in four parallel bands along the tentilla (Arrows). **A.** Viewed from the correct angle, they elongate inwards and are all oriented in the same direction

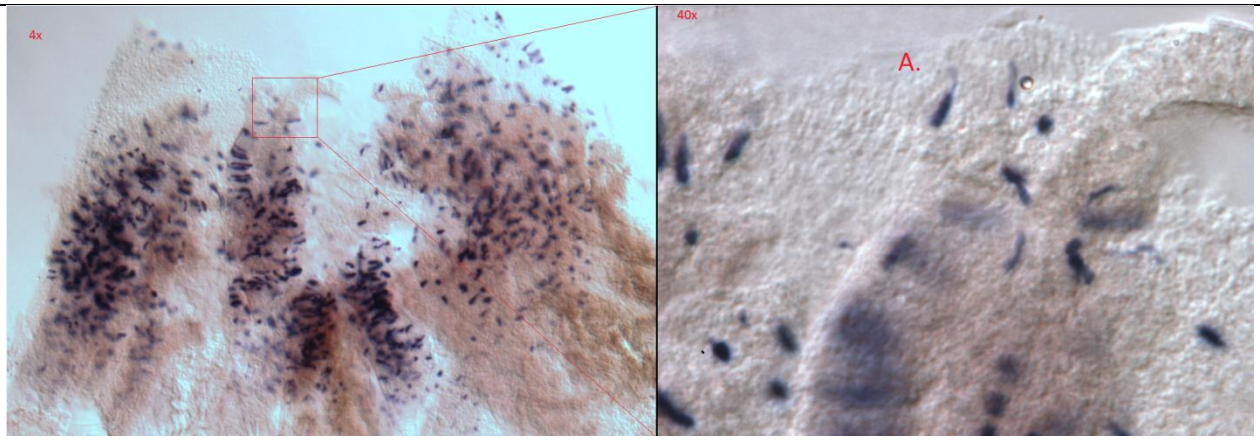

Figure 10. NP33 expression around the mouth. The region is dense with elongated NP33-expressing cells. **A.** These cells project a long, thin tentative axon structure

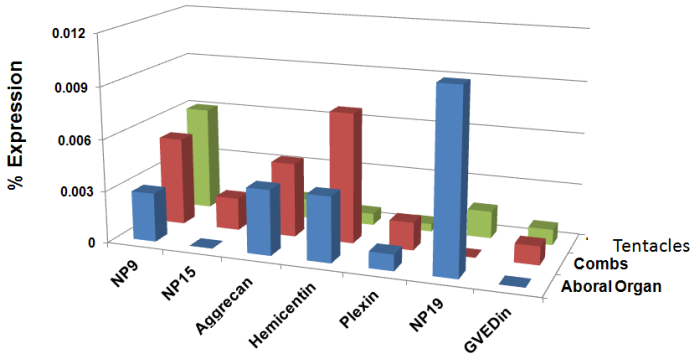

Figure 11. Expression comparison of selected peptides and common extracellular components. Hemicentin is an immunoglobulin, Aggrecan forms an extracellular matrix, while Plexin is a growth receptor associated with neuronal development. The predicted peptides show differential expression comparable to other extracellular products.

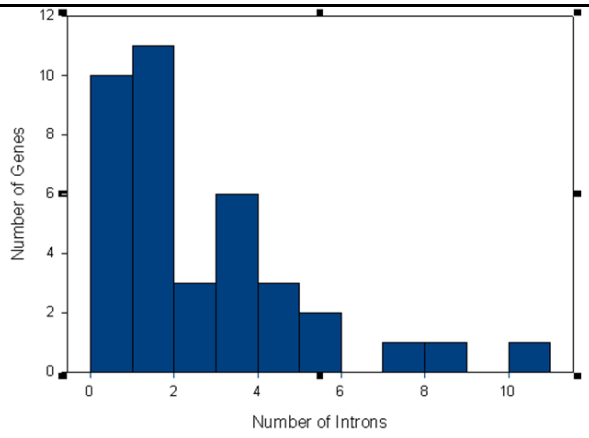

Figure 12. Intron numbers for predicted secretory peptides. The average number is very low, typically zero or one.

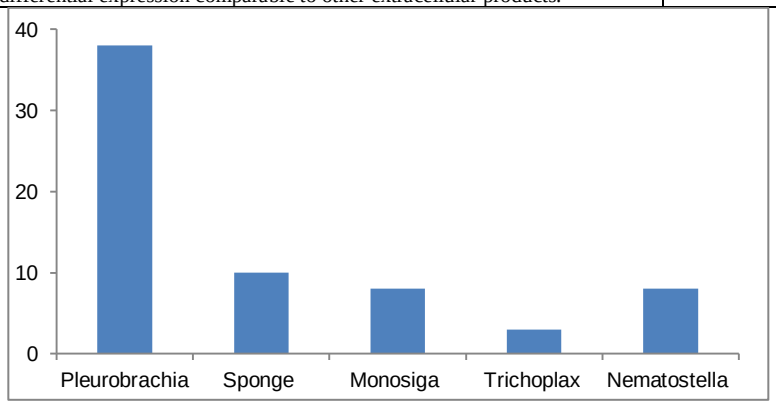

Figure 13. Predicted secretory peptide homology across 'basal' phyla. A maximum of 25% are present in the closest phylum Porifera, with significant gene loss in placazoans

Prediction Stringency	SignalP NN	SignalP HMM	TargetP	Expected of cleavage sites	Phobius TM	Phobius SP	# of predictions
Least	90%>x≥50%	90%>x≥50%	X=1	x>0	No	Yes	196
Medium	90%>x≥50%	90%>x≥50%	X=1	x>1	No	Yes	264
Most	x≥90%	x≥90%	X=1	x>1	No	Yes	38

Table1. SignalP NN and HMM are the probabilities produced by the SignalP Neural Network and Hidden Markov model Respectively. TargetP is the certainty score (1-5) produced by TargetP. Phobius TM is Phobius' prediction of a transmembrane domain. Phobius SP is Phobius' prediction of a signal peptide. Finally, the number of proteins matching these prediction criteria.

Appendix 1 – *In situ* Hybridization /Mounting Protocol

Day 1

Fix whole specimen in 4% paraformaldehyde in Filtered Sea Water (FSW) overnight (O/N) at 4^0C

Place no more than 10 animals in a 50 ml conical tube. To mix, hold on side and rotate gently.

4% Formaldehyde in Filtered Sea Water (5.4mL Formaldehyde in 44.6 ml FSW)

Fresh sea water is obtained from our tanks then filtered on 0.2 um pore membrane filter unit.

Day2

Rinse 3 x for 10 min in PTW (PBST) at Room Temperature

To mix, hold on side and rotate gently. Dispose of all solutions in hazardous waste.

PTW (PBST) 50ul Tween 20 in 50mL 1 x PBS (1 x PBS is 5 ml of 10x PBS in 45ml MQ H_2O)

Wash in 1:1 Methanol (MeOH)/PTW (to equilibrate to MeOH) 10 min at Room Temperature

1:1 MeOH/PTW (25 ml Methanol and 25 ml PTW)

Store in 100% MeOH at -20C for 2 hours up to a week

Place on side in freezer to allow animals to be separated

Day 3

Rehydrate specimen for 10 minutes in MeOH/PTW 3:1, 1:1, 1:3, 0:1 at Room Temperature

30mL/20mL/10mL MeOH fill to 40 ml with MQ H_2O

Wash in 1:1 solution of hybridization buffer (HB) and PTW for 15 minutes at Room Temperature

25ml PTW in 25 ml HB buffer

Incubate (prehybridize) in pre-HB buffer for 1 hours at 60^0C

*Hybridization buffer (HB) (50% formamide, 5mM EDTA, 5X SSC, 1X Denhardt solution (in 1.5 ml tubes at –20^0C) (0.02% ficoll, 0.02% polyvynilpirrolidon, 0.02% BSA), 0.1% Tween 20, 0.5 mg / ml yeast RNA (in–20^0C (Invitrogen) = **(for 50 ml)** 25 ml formamide, 0.5 ml 0.5M EDTA, 12.5 ml 20X SSC, 50 ul Tween20, 1 ml 50X Denhardt s-n, 25 mg yeast RNA)*

NOTE: pre-HB buffer does not contain Denhart or tRNA

Incubate (hybridize) in HB with DIG-RNA probe O/N at 60^0C

Add Denhardt and yeast tRNA to pre-HB buffer to make HB buffer. Mix 1 ml of hybridization buffer (HB) with 2-10 ul (200-400 ng Qubit) of probe. Then remove prehybridization buffer from tube with animals and add mixed hybridization buffer. Very gently rock O/N.

Day 4

Wash in pre-HB for 30 min at 60^0C

Remove old HB buffer and replace with 1 ml fresh pre-HB in same well

Wash in 1:1 HB/PTW for 30 min at 60^0C

Remove old HB buffer and replace with 1 ml 1:1 pre-HB/PTW in same well

Wash in PTW for 30 min at Room Temperature

Remove old 1:1 pre-HB/PTW and replace with 1ml PTW in same well

Block in 10% Goat Serum (GS) for 60 min at Room Temperature

Remove old PTW and replace with 1ml 10% Goat Serum in same well

1mL GS (in 1.5 ml tubes at –20^{0}C) in 10mL PBT

Incubate in anti-DIG 1/2000 at 4^{0}C O/N

Remove 10% Goat Serum. Replace with 1 ml 1% GS + 1:2000 (0.5 ul / 1 ml 1% GS) of alkaline phosphatase - conjugated DIG-antibodies in same well

0.1mL GS (in 1.5 ml tubes at –20^{0}C) in 10mL PBT

1:2000 (0.5 ul alkaline phosphatase -conjugated DIG-Antibody/ 1 ml 1% GS)

Day 5

Wash 4 x 30 min in PBS Room Temperature

Prepare 24 well plates label wells with marker the names of probes.

Make detection buffer and aliquot 1mL into clean well for each sample. When ready to develop, add 20uL of NBT/BICP mix until dissolved. Should be yellow in color! **NOW** add samples. **Put on ICE and cover with tin foil.**

*Detection buffer (100 mM NaCl, 50mM MgCl2, 0.1% Tween 20, 1mM levamisole, 100mM Tris HCl = (**for 50 ml**) 1 ml of 5 M NaCl, 2.5ml of 1 M MgCl(10g/50mlH_2O), 0.012g levamisole, 50ul Tween 20, 5ml of 1M Tris-HCl pH=8; adjust PH to 9.5 with 10M NaOH, filter the resulting solution)*

Do not keep detection buffer longer than 2 weeks or if it becomes cloudy.

Watch for appropriate color development

If crystals begin pooling non-specifically in the mesoglea, cut animal in half and shake gently to evacuate interior

Stop in 4% paraformaldehyde in MeOH

Wash in 4% paraformaldehyde in MeOH 30 min Room Temperature

Timing of this step depends on the strength of the signal and the background. This time is for high signal with low background

Change solution couple times

5.4mL Formaldehyde in 44.6 ml Methanol

Wash 3x 10 min in Ethanol (EtOH) Room Temperature

Begin mounting immediately, or store in 100% EtOH for up to 4 days at 4^{o}C

Specific staining begins to fade after 4 days

***In situ* mounting protocol**

1. Arrange 3 wells of methyl salicylate
2. Move sample to well plate with wells for PBS, 4% paraformaldehyde in MeOH, 100% EtOH for each sample. PBS first
3. When all materials are assembled, transfer to para/MeOH. Let sit for a few minutes
4. Perform isolating cuts in the MeOH, as EtOH causes sample to shrink and become brittle
5. Transfer to EtOH. Stretch sample gently to counter shriveling
6. Transfer to methyl salicylate.
7. Wait until the sample begins to sink before transferring to the next methyl well. Repeat for the third well
8. Clean slide of dust with kim wipe
9. Transfer sample to slide with forceps
10. Drip permount onto sample so as to avoid bubbles (3 drops for thin samples)
11. Lower cover slip onto sample diagonally, so as to avoid bubbles
12. Gently press cover slip to spread permount. Add more via capillary action if necessary.

Citations

Bruce R. Southey, A. A., Tyler A. Zimmerman, Sandra L. Rodriguez-Zas, and Jonathan V. Sweedler (2006). "NeuroPred: a tool to predict cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides." Nucleic Acids Research **34**: W267-W272.

Hawkins, J. a. B., M. (2005). "Detecting and Sorting Targeting Peptides with Recurrent Networks and Support Vector Machines." Journal of Bioinformatics and Computational Biology **4**(1): 1-18.

Jannick Dyrløv Bendtsen, H. N., Gunnar von Heijne and Søren Brunak (2004). "Improved prediction of signal peptides: SignalP 3.0." Journal of Molecular Biology **340**: 783-795.

Krogh A, L. B., von Heijne G, Sonnhammer EL. (2001). "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes." Journal of Molecular Biology **305**: 567-580.

Lukas Käll, A. K. a. E. L. L. S. (2004). "A Combined Transmembrane Topology and Signal Peptide Prediction Method." Journal of Molecular Biology **338**: 1024-1036.

Olof Emanuelsson, H. N., Søren Brunak and Gunnar von Heijne (2000). "Predicting subcellular localization of proteins based on their N-terminal amino acid sequence." Journal of Molecular Biology **300**: 1005-1016.

Mathew Citarella, David Girardo, Andrea Kohn, Leonid Moroz (2012). "Global Discovery and Validation of Signaling Molecules in the Ctenophore" SICB Conference P1.135

David Girardo, Mathew Citarella, Andrea Kohn, Leonid Moroz (2012) "Automatic transcriptome analysis and quest for signaling molecules in basal metazoans" SICB Conference P1.136

Mathew Citarella, David Girardo, Joshua Swore, Andrea Kohn, Leonid Moroz (2011). "Genome Wide Analysis of Neurotransmitter Signalling in the Ctenophore, Pleurobrachia Bachei" 12th Symposium on Invertebrate Neurobiology, Tihany, Hungary 15

Hajj-Ali I,Anctil M. (1997) "Characterization of serotonin receptors in cnidarian *Renilla koellikeri* : a radiobonding analysis" Neurochem 83-93

Leonid L. Moroz (2009) "On the Independent Origins of Complex Brains and Neurons" Brain Behavior and Evolution **77**:177-90

Swore JJ, Kohn AB, Citarella MR, Bobkova YV, Moroz LL (2012) "Molecular mapping of ctenophore neurons and glutamate singaling" SICB Conference P2.132

Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T,Kuo A, Mitros T, Salamov A, Carpenter ML, Signorovitch AY, Moreno MA,Kamm K, Grimwood J, Schmutz J, Shapiro H, Grigoriev IV, Buss LW,Schierwater B, Dellaporta SL, Rokhsar DS (2008) "The Trichoplax genome and the nature of placozoans" Nature **454:**955-960

Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier ME, Mitros T,Richards GS, Conaco C, Dacre M, Hellsten U, Larroux C, Putnam NH,Stanke M, Adamska M, Darling A, Degnan SM, Oakley TH, Plachetzki DC,Zhai Y, Adamski M, Calcino A, Cummins SF, Goodstein DM, Harris C,Jackson DJ, Leys SP, Shu S, Woodcroft BJ, Vervoort M, Kosik KS, Manning G,Degnan BM, Rokhsar DS (2010) "The *Amphimedon queenslandica* genome and the evolution of animal complexity" Nature **466:**720-726

Citarellla MR, Kohn AB, Bobkova E, Yu F, Moroz LL (2011) "Genome-wide Characterization of Signaling Peptides Across Animal Phyla and Parallel Evolution of Neural Systems" SICB Conference P3.97

J.J. Hopfield (1982) "Neural Networks and Physical Systems with emergent Collective Computational Abilities" Proc. NatL Acad. Sci. USA **79**:2554-2558

Mark Stamp (2004) "A revealing introduction to hidden Markov Models"

Jager, M., Chiori, R., Alié, A., Dayraud, C., Quéinnec, E. and Manuel, M. (2011), "New insights on ctenophore neural anatomy: Immunofluorescence study in *Pleurobrachia pileus"* Journal of Experimental Zoology **316B**: 171–187