Proteome-wide neuropeptide identification using NeuroPeptide-HMMer (NP-HMMer)

Meet Zandawala[1,2,3,#], Muhammad Bilal Amir[1], Joel Shin[1], Won C. Yim[1] and Luis Alfonso Yañez Guerra[4,5,#]
[1] Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV 89557, USA
[2] Integrative Neuroscience Program, University of Nevada, Reno, NV 89557, USA
[3] Neurobiology and Genetics, Theodor-Boveri-Institute, Biocenter, Julius-Maximilians-University of Würzburg, Am Hubland, 97074 Würzburg, Germany
[4] School of Biology, University of Southampton, University Road SO17 1BJ, Southampton, UK
[5] Institute for Life Sciences, University of Southampton, University Road SO17 1BJ, Southampton, UK
[#] Equal contribution. Correspondence should be addressed to M.Z. (mzandawala@unr.edu) and L.A.Y.G. (L.A.Yanez-Guerra@soton.ac.uk)
ORCID: M.Z. 0000-0001-6498-2208; W.C.Y. 0000-0002-7489-0435; L.A.Y.G. 0000-0002-2523-1310

Abstract

Neuropeptides are essential neuronal signaling molecules that orchestrate animal behavior and physiology via actions within the nervous system and on peripheral tissues. Due to the small size of biologically active mature peptides, their identification on a proteome-wide scale poses a significant challenge using existing bioinformatics tools like BLAST. To address this, we have developed NeuroPeptide-HMMer (NP-HMMer), a hidden Markov model (HMM)-based tool to facilitate neuropeptide discovery, especially in underexplored invertebrates. NP-HMMer utilizes manually curated HMMs for 46 neuropeptide families, enabling rapid and accurate identification of neuropeptides. Validation of NP-HMMer on *Drosophila melanogaster*, *Daphnia pulex*, *Tribolium castaneum* and *Tenebrio molitor* demonstrated its effectiveness in identifying known neuropeptides across diverse arthropods. Additionally, we showcase the utility of NP-HMMer by discovering novel neuropeptides in Priapulida and Rotifera, identifying 22 and 19 new peptides, respectively. This tool represents a significant advancement in neuropeptide research, offering a robust method for annotating neuropeptides across diverse proteomes and providing insights into the evolutionary conservation of neuropeptide signaling pathways.

Keywords: Neuropeptide; Hidden Markov model; Invertebrate; Evolution; Rotifer; Priapulid

Introduction

Neuropeptides and peptide hormones are the largest class of neurochemical messengers. They regulate diverse animal behaviors and physiological processes including feeding, locomotion, metabolism, growth, and reproduction (Nässel and Zandawala, 2019). Consequently, they are implicated in various diseases ranging from obesity and diabetes to high blood pressure and sleep disorders. Further, the origins of several neuropeptide families can be traced back to the common ancestor of Bilateria (Jekely, 2013; Mirabeau and Joly, 2013). Thus, orthologs of several vertebrate neuropeptide families are found in invertebrates such as insects, nematode and polychaete worms, and echinoderms. These include orthologs of vertebrate neuropeptides like calcitonin, galanin, orexin, neuropeptide-S, gonadotropin-releasing hormone (GnRH), thyrotropin-releasing hormone (TRH), corticotropin-releasing factor, glycoprotein hormones, insulin, substance P, neuropeptide Y, and vasopressin/oxytocin, amongst others (Bauknecht and Jekely, 2015; Istiban et al., 2024; Lindemans et al., 2009; Nässel and Zandawala, 2019; Nässel et al., 2019; Odekunle et al., 2019; Semmens et al., 2015; Stafflinger et al., 2008; Staubli et al., 2002; Tian et al., 2016; Van Sinay et al., 2017; Wegener and Chen, 2022; Zandawala, 2012). In many instances, the functions of these signaling systems have also been conserved throughout evolution (Nässel and Zandawala, 2019; Nässel et al., 2019; Van Sinay et al., 2017). Therefore, discovering and characterizing neuropeptide signaling pathways in invertebrates can offer valuable insights into their vertebrate orthologs.

Despite their widespread occurrence and medical significance, establishing evolutionary relationships between neuropeptides across different animal phyla has historically been quite challenging due to the short sequences of biologically active mature peptides. Additionally, bioinformatics approaches such as Basic Local Alignment Search Tool (BLAST) are often unsuitable for discovering neuropeptide precursors in animals belonging to phyla other than Arthropoda, Nematoda, Annelida, Mollusca, and Chordata, where many neuropeptides have already been identified. This limitation is particularly pronounced for neuropeptide precursors that generate mature peptides with limited conservation beyond the phyla in which they were originally discovered. For instance, although the structure of vertebrate TRH (comprised of only 3 amino acids) was discovered in 1969 (Boler et al., 1969; Burgus et al., 1969), it was not until 2015 that the identity of invertebrate TRH was revealed through the functional characterization of *Platynereis dumerilii* TRH receptors (Bauknecht and Jekely, 2015). While this issue can be circumvented by using neuropeptide prediction tools based on machine learning models (Agrawal et al., 2019; Kang et al., 2019; Ofer and Linial, 2014; Wang et al., 2023; Wang et al., 2024), these existing tools are unable to determine the neuropeptide family to which the predicted neuropeptide precursor belongs. The large number of neuropeptide families, some of which have similar conserved motifs, further complicates accurate classification of neuropeptides.

To address this gap, we have developed NeuroPeptide-HMMer (NP-HMMer), a hidden Markov model (HMM)-based tool designed to facilitate neuropeptide discovery, particularly in understudied arthropods and other invertebrate phyla. NP-HMMer is based on manually curated HMMs for 46 invertebrate neuropeptides, which enables rapid and accurate identification of neuropeptides in entire proteomes by presenting output in multiple forms (i.e., trimmed and non-trimmed sequence alignments, and sequence logos). We demonstrate the effectiveness of this tool by discovering members of several neuropeptide families in Priapulida (penis worms) and Rotifera (wheel animalcules).

Methods

Identification and curation of neuropeptide precursors for HMM generation

Members of 46 neuropeptide precursor families were used as queries to search for orthologs from proteomes of the following 17 representative arthropod species: *Apis mellifera, Aedes aegypti, Bombyx mori, Acyrthosiphon pisum, Daphnia pulex, Drosophila melanogaster, Ixodes scapularis, Lepeophtheirus salmonis, Nasonia vitripennis, Locusta migratoria, Pediculus humanus, Rhodnius*

prolixus, Strigamia maritima, Tetranychus urticae, Tribolium castaneum, Varroa destructor, and Zootermopsis nevadensis. These species encompass a broad phylogenetic range that includes insects, crustaceans, myriapods, and chelicerates. Neuropeptide precursors were identified using BLASTp with an E-value threshold of 1e-2 to ensure the identification of orthologs. The search was limited to the top three hits for each query sequence and the top high-scoring pair for each target sequence. Following BLASTp, the sequences were manually curated to remove redundant or partial sequences. Sequences not encoding neuropeptides were also removed by manually scanning for characteristic prohormone cleavage sites and conserved motifs in mature peptides. Signal peptides were identified using SignalP 6.0 (Teufel et al., 2022).

NP-HMMer generation and validation

Ortholog sequences belonging to the same neuropeptide family were aligned using MUSCLE (Edgar, 2004), followed by manual refinement to optimize the positioning of conserved residues and motifs. HMMER software suite (http://hmmer.org/) was then used to construct and train family-specific HMMs using the refined sequence alignments. HMMs were trained using default parameters and the hmmbuild command, with a minimum of five neuropeptide precursor sequences used for each model. Proteomes of *Drosophila melanogaster*, *Daphnia pulex* and *Tribolium castaneum* were scanned using the trained HMM profiles with hmmsearch and an E-value of 1e-2. Identified sequences were manually inspected to confirm characteristic neuropeptide features for each family. In most of the cases, the models identified the correct neuropeptide. In cases where the model did not obtain accurate hits, sequence alignments were modified, and models were generated.

Neuropeptide identification in Priapulida and Rotifera

Validated models were used to identify neuropeptide precursors in whole predicted proteomes of one priapulid (*Priapulus caudatus*, NCBI; GCF_000485595.1) and four rotifers ((*Adineta ricciae* (UNIPROT; UP000663828), *Brachionus plicatilis* (UNIPROT UP000276133), *Didymodactylos carnosus* (UNIPROT; UP000663829) and *Rotaria socialis* (UP000663873)) using an E-value of 1e-2, and -domE e1-6.

Sequence annotation and visualization

Signal peptides were predicted using SignalP 6.0. Monobasic and dibasic cleavage sites, and mature peptides were manually annotated based on the NP-HMMer output. Sequence alignments in the manuscript were generated using Clustal Omega (https://www.ebi.ac.uk/jdispatcher/msa/clustalo) with default settings. The alignments were shaded using Boxshade (https://junli.netlify.app/apps/boxshade/) based on at least 60% amino acid conservation as the minimum for highlighting.

Results and discussion

NP-HMMer generation

HMM-based searches can be powerful in discovering orthologs of genes that evade BLAST-based searches. This is due to their ability to model positional dependencies between amino acids, allowing HMMs to capture conserved motifs interrupted by insertions or deletions. HMMs use a probabilistic framework to represent amino acid distributions, accommodating variability within a family while still detecting distant homologs. Furthermore, HMMs can be used to construct profile models that represent the consensus of a protein family, enabling the identification of new family members with limited sequence similarity to known members. We previously used this approach to identify GnRH in the amphioxus *Branchiostoma floridae* (Yanez Guerra and Zandawala, 2023). In contrast, BLAST, relying on pairwise alignments, may struggle to detect relationships when sequence similarity is low (Eddy, 2004). Hence, we have generated NP-HMMer **(Figure 1)**, a tool that allows annotation of 46 invertebrate neuropeptides in whole proteomes. NP-HMMer is based on python and the open-source version is freely-available with step-by-step instructions for setup and execution locally.

NP-HMMer validation

We validated our models by testing their effectiveness in identifying neuropeptides in proteomes of *Drosophila melanogaster* **(Figure S1)**, *Daphnia plex* **(Figure S2)** and *Tribolium castaneum* **(Figure S3)**, whose neuropeptidomes have been previously established (Dircksen et al., 2011; Hewes and Taghert, 2001; Li et al., 2008; Nässel and Zandawala, 2019). Our search identified members of all neuropeptide families in these species **(Table S1)**. *Daphnia* diuretic hormone 44 (DH44) and myosuppressin, as well as *Tribolium* natalisin were missing in the proteomes used for the analysis. Hence, manual addition of these neuropeptide sequences to their respective proteomes enabled their detection by NP-HMMer. Since sequences from *Drosophila, Daphnia* and *Tribolium* were used to generate the models, we also tested the effectiveness of NP-HMMer using a proteome from *Tenebrio mollitor* whose sequences were not used to generate NP-HMMer. We were able to retrieve all the neuropeptides previously identified in *Tenebrio* (Veenstra, 2019) for which the models are available **(Table S2)**. In addition, we also identified additional neuropeptide isoforms and transcript variants not identified previously. Hence, NP-HMMEr is effective at identifying neuropeptides in a wide-range of insects.

In some cases, our models could not distinguish between closely-related neuropeptides. For example, allatostatin-C, allatostatin-CC, and allatostatin-CCC have similar signatures that prevent automatic classification (Veenstra, 2016). Similarly, adipokinetic hormone (AKH), corazonin and AKH/corazonin-related peptide (ACP), bursicon alpha and bursicon beta, CCHamide-1 and CCHamide-2, as well as natalisin and tachykinin are quite similar which prevents unambiguous identification. However, manual examination of the search results, which are conveniently presented in the form of sequence alignments and logos **(Figure S4)**, can readily resolve any discrepancies. We also provide a reference guide to facilitate the identification of these closely-related neuropeptides **(Figure S5)**. Regardless, this approach is still effective as we were able to identify phoenixin precursor **(Figure S1)** for the first time in *Drosophila melanogaster.*

Neuropeptide identification in Priapulida and Rotifera

Having validated NP-HMMer, we decided to comprehensively identify neuropeptide complements of Priapulida and Rotifera. Previous studies have independently identified orthologs of AKH-like (Li et al., 2016), neuropeptide F (Yanez-Guerra et al., 2020), RYamide/Luqin (Yanez-Guerra et al., 2018), pigment-dispersing factor (Mayer et al., 2015), ion transport peptide (Gera et al., 2024), vasopressin/oxytocin (VP/OXT) (Lockard et al., 2017), allatostatin-CC (misclassified as allatostatin-C) and FMRFamide-like peptides (Christie et al., 2011) in priapulids. In contrast, to the best of our knowledge, only AKH-like precursor has been discovered in rotifers (Cadena-Caballero et al., 2023; Hauser and Grimmelikhuijzen, 2014). Hence, we chose species from these two phyla to assess the suitability of NP-HMMer in discovering orthologs of arthropod neuropeptides in other phyla. Specifically, we mined proteomes of one priapulid (*Priapulus caudatus* **(Figure S6)**) and four rotifers (*Adineta ricciae* **(Figure S7)**, *Brachionus plicatilis* **(Figure S8)**, *Didymodactylos carnosus* **(Figure S9)** and *Rotaria socialis* **(Figure S10)**) using NP-HMMer. Our search identified members of 25 and 20 neuropeptide families in Priapulida and Rotifera, respectively (**Table 1, Figures 2-3 and S11**). When compared to previously discovered neuropeptides in these phyla, this represents 22 and 19 novel neuropeptides in Priapulida and Rotifera, respectively. In particular, we identified orthologs of crustacean cardioactive peptide (CCAP) **(Figure 2B)**, leucokinin **(Figure 2C)**, allatostatin-A **(Figure 2F)**, proctolin **(Figure 2H)**, allatostatin-CCC **(Figure 2I)**, pigment-dispersing factor (PDF) **(Figure 2L)**, VP/OXT **(Figure 2N)** FMRFamide **(Figure 3C)**, CCHamide **(Figure 3D)**, CAPA **(Figure 3E)**, neuropeptide F **(Figure 3H)**, orcokinin **(Figure 3I)**, phoenixin **(Figure 3J)**, bursicon alpha, glycoprotein hormone alpha 2 (GPA2) and insulin-like peptides **(Figure S11)** for the first time in rotifers**.** We also identified precursors encoding two copies of Wamides in *Adineta* **(Figure S7)** and *Didymodactylos* **(Figure S9)**. These peptides appear to be the ancestral form of allatostatin-B as loss of a cleavage site in between the two

Adineta Wamides could give rise to an allatostatin-B-like peptide **(Figure 3F)**. Moreover, we identified a novel AKH-like precursor in *Brachionus.* Sequence alignment suggests that this AKH is more similar to other arthropod AKH peptides, whereas the AKH-like peptide discovered previously (Hauser and Grimmelikhuijzen, 2014) appears to be more similar to arthropod ACP **(Figure 3G)**. Functional characterization of AKH and ACP receptors in the future is needed to unambiguously determine their ligands. This limitation also holds true for all other neuropeptides identified here as they can only be considered predictions until functionally characterized. Lastly, we also identified a calcitonin ortholog in *Priapulus* which is significantly shorter compared to other arthropod calcitonin **(Figure 3A)**. Additional priapulid transcriptomes/genomes need to be examined to ensure that the truncation is not an artifact caused during sequence assembly. The inability to detect more neuropeptides from these datasets could be attributed to query proteomes not enriched from neural tissue. Nonetheless, our analyses showcase the power of NP-HMMer by rapidly and reliably discovering neuropeptides in diverse protostomes. One limitation of the models generated here is that they are largely not suitable for discovering neuropeptides in non-bilaterians or deuterostomes. Therefore, we plan to expand NP-HMMer to include models for additional neuropeptides, including those that are predominantly found in non-protostomian invertebrates.

Acknowledgments

We would like to thank Dr. Ismail Moghul and Dr. Maurice Elphick for the initial discussions and Dr. Theresa McKim for helpful feedback during the preparation of this manuscript. L.A.Y.G. was supported by a BBSRC fellowship (BB/W010305/1) and funding from the Royal Society (RG\R1\241397).

Author contributions

Study conception: L.A.Y.G. and M.Z.; Computational analyses: L.A.Y.G., J.S. and W.Y.; Data analyses: M.B.A., L.A.Y.G. and M.Z.; Data visualization: all authors; Manuscript writing: M.Z.; Manuscript editing: all authors.

Declaration of competing interest

We declare that we have no competing interests.

Code availability

All the models, code, and files to automatically create the alignments are freely available on Github (https://github.com/Imnotabioinformatician/NP-HMMer)

References

Agrawal, P., Kumar, S., Singh, A., Raghava, G.P.S., Singh, I.K., 2019. NeuroPIpred: a tool to predict, design and scan insect neuropeptides. Sci Rep 9(1), 5129.

Bauknecht, P., Jekely, G., 2015. Large-Scale Combinatorial Deorphanization of *Platynereis* Neuropeptide GPCRs. Cell Rep 12(4), 684-693.

Boler, J., Enzmann, F., Folkers, K., Bowers, C.Y., Schally, A.V., 1969. The identity of chemical and hormonal properties of the thyrotropin releasing hormone and pyroglutamyl-histidyl-proline amide. Biochem Biophys Res Commun 37(4), 705-710.

Burgus, R., Dunn, T.F., Desiderio, D., Guillemin, R., 1969. Molecular structure of the hypothalamic hypophysiotropic TRF factor of ovine origin: mass spectrometry demonstration of the PCA-His-Pro-NH2 sequence. C R Acad Hebd Seances Acad Sci D 269(19), 1870-1873.

Cadena-Caballero, C.E., Munive-Arguelles, N., Vera-Cala, L.M., Barrios-Hernandez, C., Duarte-Bernal, R.O., Ayus-Ortiz, V.L., Pardo-Diaz, L.A., Agudelo-Rodriguez, M., Bautista-Rozo, L.X., Jimenez-Gutierrez, L.R., Martinez-Perez, F., 2023. APGW/AKH Precursor from Rotifer *Brachionus plicatilis* and the DNA Loss Model Explain Evolutionary Trends of the Neuropeptide LWamide, APGWamide, RPCH, AKH, ACP, CRZ, and GnRH Families. J Mol Evol 91(6), 882-896.

Christie, A.E., Nolan, D.H., Garcia, Z.A., McCoole, M.D., Harmon, S.M., Congdon-Jones, B., Ohno, P., Hartline, N., Congdon, C.B., Baer, K.N., Lenz, P.H., 2011. Bioinformatic prediction of arthropod/nematode-like peptides in non-arthropod, non-nematode members of the Ecdysozoa. Gen Comp Endocrinol 170(3), 480-486.

Dircksen, H., Neupert, S., Predel, R., Verleyen, P., Huybrechts, J., Strauss, J., Hauser, F., Stafflinger, E., Schneider, M., Pauwels, K., Schoofs, L., Grimmelikhuijzen, C.J., 2011. Genomics, transcriptomics, and peptidomics of *Daphnia pulex* neuropeptides and protein hormones. J Proteome Res 10(10), 4478-4504.

Eddy, S.R., 2004. Where did the BLOSUM62 alignment score matrix come from? Nat Biotechnol 22(8), 1035-1036.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5), 1792-1797.

Gera, J., Agard, M., Nave, H., Sajadi, F., Thorat, L., Kondo, S., Nässel, D.R., Paluzzi, J.-P.V., Zandawala, M., 2024. Anti-diuretic hormone ITP signals via a guanylate cyclase receptor to modulate systemic homeostasis in *Drosophila*. Cold Spring Harbor Laboratory.

Hauser, F., Grimmelikhuijzen, C.J., 2014. Evolution of the AKH/corazonin/ACP/GnRH receptor superfamily and their ligands in the Protostomia. Gen Comp Endocrinol 209, 35-49.

Hewes, R.S., Taghert, P.H., 2001. Neuropeptides and neuropeptide receptors in the *Drosophila melanogaster* genome. Genome Res 11(6), 1126-1142.

Istiban, M.N., De Fruyt, N., Kenis, S., Beets, I., 2024. Evolutionary conserved peptide and glycoprotein hormone-like neuroendocrine systems in C. elegans. Mol Cell Endocrinol 584, 112162.

Jekely, G., 2013. Global view of the evolution and diversity of metazoan neuropeptide signaling. Proc Natl Acad Sci U S A 110(21), 8702-8707.

Kang, J., Fang, Y., Yao, P., Li, N., Tang, Q., Huang, J., 2019. NeuroPP: A Tool for the Prediction of Neuropeptide Precursors Based on Optimal Sequence Composition. Interdiscip Sci 11(1), 108-114.

Li, B., Predel, R., Neupert, S., Hauser, F., Tanaka, Y., Cazzamali, G., Williamson, M., Arakane, Y., Verleyen, P., Schoofs, L., Schachtner, J., Grimmelikhuijzen, C.J., Park, Y., 2008. Genomics, transcriptomics, and peptidomics of neuropeptides and protein hormones in the red flour beetle *Tribolium castaneum*. Genome Res 18(1), 113-122.

Li, S., Hauser, F., Skadborg, S.K., Nielsen, S.V., Kirketerp-Moller, N., Grimmelikhuijzen, C.J., 2016. Adipokinetic hormones and their G protein-coupled receptors emerged in Lophotrochozoa. Sci Rep 6, 32789.

Lindemans, M., Liu, F., Janssen, T., Husson, S.J., Mertens, I., Gade, G., Schoofs, L., 2009. Adipokinetic hormone signaling through the gonadotropin-releasing hormone receptor modulates egg-laying in *Caenorhabditis elegans*. Proc Natl Acad Sci U S A 106(5), 1642-1647.

Lockard, M.A., Ebert, M.S., Bargmann, C.I., 2017. Oxytocin mediated behavior in invertebrates: An evolutionary perspective. Dev Neurobiol 77(2), 128-142.

Mayer, G., Hering, L., Stosch, J.M., Stevenson, P.A., Dircksen, H., 2015. Evolution of pigment-dispersing factor neuropeptides in Panarthropoda: Insights from Onychophora (velvet worms) and Tardigrada (water bears). J Comp Neurol 523(13), 1865-1885.

Mirabeau, O., Joly, J.S., 2013. Molecular evolution of peptidergic signaling systems in bilaterians. Proc Natl Acad Sci U S A 110(22), E2028-2037.

Nässel, D.R., Zandawala, M., 2019. Recent advances in neuropeptide signaling in *Drosophila*, from genes to physiology and behavior. Prog Neurobiol 179, 101607.

Nässel, D.R., Zandawala, M., Kawada, T., Satake, H., 2019. Tachykinins: Neuropeptides That Are Ancient, Diverse, Widespread and Functionally Pleiotropic. Front Neurosci 13, 1262.

Odekunle, E.A., Semmens, D.C., Martynyuk, N., Tinoco, A.B., Garewal, A.K., Patel, R.R., Blowes, L.M., Zandawala, M., Delroisse, J., Slade, S.E., Scrivens, J.H., Egertova, M., Elphick, M.R., 2019. Ancient role of vasopressin/oxytocin-type neuropeptides as regulators of feeding revealed in an echinoderm. BMC Biol 17(1), 60.

Ofer, D., Linial, M., 2014. NeuroPID: a predictor for identifying neuropeptide precursors from metazoan proteomes. Bioinformatics 30(7), 931-940.

Semmens, D.C., Beets, I., Rowe, M.L., Blowes, L.M., Oliveri, P., Elphick, M.R., 2015. Discovery of sea urchin NGFFFamide receptor unites a bilaterian neuropeptide family. Open Biol 5(4), 150030.

Stafflinger, E., Hansen, K.K., Hauser, F., Schneider, M., Cazzamali, G., Williamson, M., Grimmelikhuijzen, C.J., 2008. Cloning and identification of an oxytocin/vasopressin-like receptor and its ligand from insects. Proc Natl Acad Sci U S A 105(9), 3262-3267.

Staubli, F., Jorgensen, T.J., Cazzamali, G., Williamson, M., Lenz, C., Sondergaard, L., Roepstorff, P., Grimmelikhuijzen, C.J., 2002. Molecular identification of the insect adipokinetic hormone receptors. Proc Natl Acad Sci U S A 99(6), 3446-3451.

Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gislason, M.H., Pihl, S.I., Tsirigos, K.D., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 40(7), 1023-1025.

Tian, S., Zandawala, M., Beets, I., Baytemur, E., Slade, S.E., Scrivens, J.H., Elphick, M.R., 2016. Urbilaterian origin of paralogous GnRH and corazonin neuropeptide signalling pathways. Sci Rep 6, 28788.

Van Sinay, E., Mirabeau, O., Depuydt, G., Van Hiel, M.B., Peymen, K., Watteyne, J., Zels, S., Schoofs, L., Beets, I., 2017. Evolutionarily conserved TRH neuropeptide pathway regulates growth in *Caenorhabditis elegans*. Proc Natl Acad Sci U S A 114(20), E4065-E4074.

Veenstra, J.A., 2016. Allatostatins C, double C and triple C, the result of a local gene triplication in an ancestral arthropod. Gen Comp Endocrinol 230-231, 153-157.

Veenstra, J.A., 2019. Coleoptera genome and transcriptome sequences reveal numerous differences in neuropeptide signaling between species. PeerJ 7, e7144.

Wang, L., Huang, C., Wang, M., Xue, Z., Wang, Y., 2023. NeuroPred-PLM: an interpretable and robust model for neuropeptide prediction by protein language model. Brief Bioinform 24(2).

Wang, L., Zeng, Z., Xue, Z., Wang, Y., 2024. DeepNeuropePred: A robust and universal tool to predict cleavage sites from neuropeptide precursors by protein language model. Comput Struct Biotechnol J 23, 309-315.

Wegener, C., Chen, J., 2022. Allatostatin A Signalling: Progress and New Challenges From a Paradigmatic Pleiotropic Invertebrate Neuropeptide Family. Front Physiol 13, 920529.

Yanez Guerra, L.A., Zandawala, M., 2023. Discovery of Paralogous GnRH and Corazonin Signaling Systems in an Invertebrate Chordate. Genome Biol Evol 15(7).

Yanez-Guerra, L.A., Delroisse, J., Barreiro-Iglesias, A., Slade, S.E., Scrivens, J.H., Elphick, M.R., 2018. Discovery and functional characterisation of a luqin-type neuropeptide signalling system in a deuterostome. Sci Rep 8(1), 7220.

Yanez-Guerra, L.A., Zhong, X., Moghul, I., Butts, T., Zampronio, C.G., Jones, A.M., Mirabeau, O., Elphick, M.R., 2020. Echinoderms provide missing link in the evolution of PrRP/sNPF-type neuropeptide signalling. Elife 9.

Zandawala, M., 2012. Calcitonin-like diuretic hormones in insects. Insect Biochem Mol Biol 42(10), 816-825.

Figures

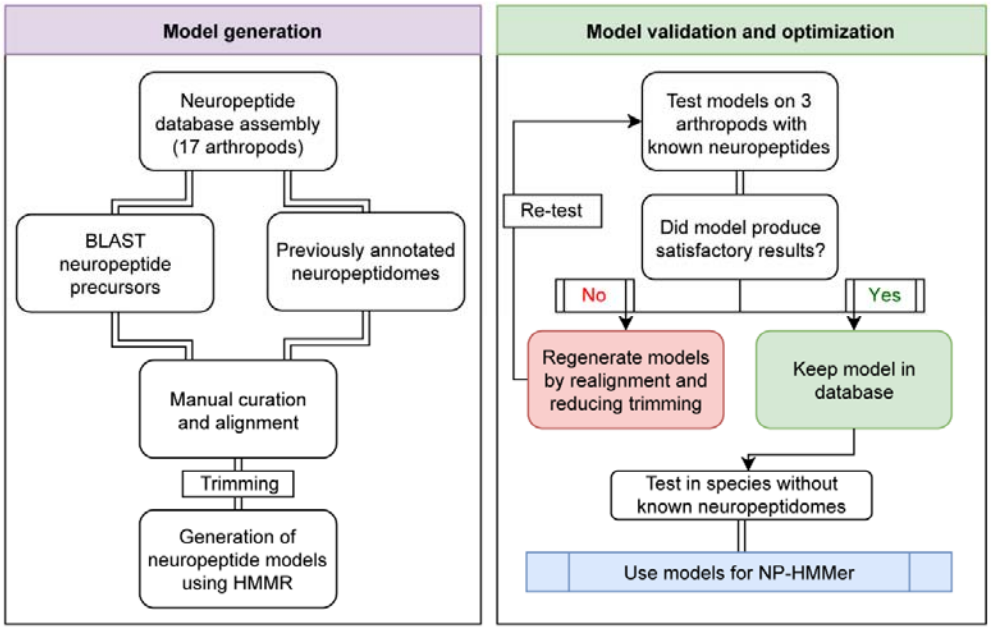

Figure 1: Workflow for creating the NP-HMMer search tool. A database of neuropeptide precursors from up to 17 representative arthropod species was first assembled. This was achieved using a combination of BLAST analyses and by manually searching previously annotated neuropeptidomes. Complete precursor sequences belonging to specific neuropeptide families were curated and aligned. The resulting alignments were trimmed to exclude non-conserved regions and HMMs were generated. The models were tested using 3 arthropod proteomes whose neuropeptidomes were previously established. Models producing accurate predictions were kept, while others were refined and retested. The final models formed the basis of NP-HMMer which was used to identify neuropeptides in priapulids and rotifers.

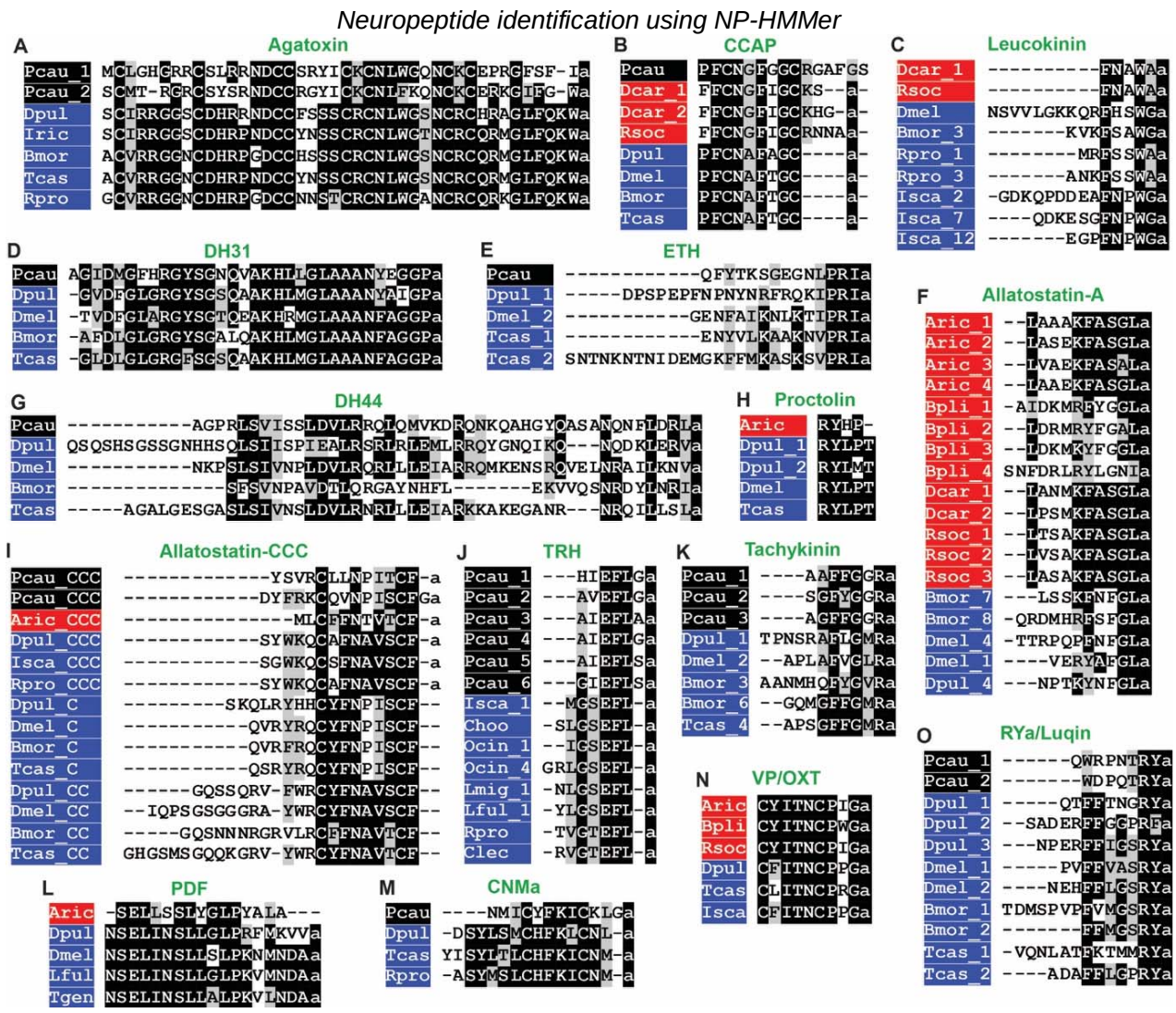

Figure 2: Multiple sequence alignments of neuropeptides discovered in Priapulida and Rotifera. Alignments of **(A)** agatoxin, **(B)** crustacean cardioactive peptide (CCAP), **(C)** leucokinin, **(D)** diuretic hormone 31 (DH31), **(E)** ecdysis-triggering hormone (ETH), **(F)** allatostatin-A, **(G)** diuretic hormone 44 (DH44), **(H)** proctolin, **(I)** allatostatin-CCC, **(J)** thyrotropin-releasing hormone (TRH), **(K)** tachykinin, **(L)** pigment-dispersing factor (PDF), **(M)** CNMamide (CNMa), **(N)** vasopressin/oxytocin (VP/OXT) and **(O)** RYamide (RYa)/Luqin mature peptides. Priapulid species are colored in black, rotifers in red, and arthropods in blue. Conserved residues are highlighted in black or gray. Species names: Pcau, *Priapulis caudatus*; Dcar, *Didymodactylos carnosus*; Rsoc, *Rotaria socialis*; Aric, *Adineta ricciae*; Bpli, *Brachionus plicatilis*; Dpul, *Daphnia pulex*; Iric, *Ixodes Ricinus*; Bmor, *Bombyx mori*; Tcas, *Tribolium castaneum*; Rpro, *Rhodnius prolixus*; Dmel, *Drosophila melanogaster*; Isca, *Ixodes scapularis*; Lful, *Ladona fulva*; Tgen, *Timema genevievae*; Choo, *Clitarchus hookeri*; Ocin, *Orchesella cincta*; Lmig, *Locusta migratoria*; Clec, *Cimex lectularius*.

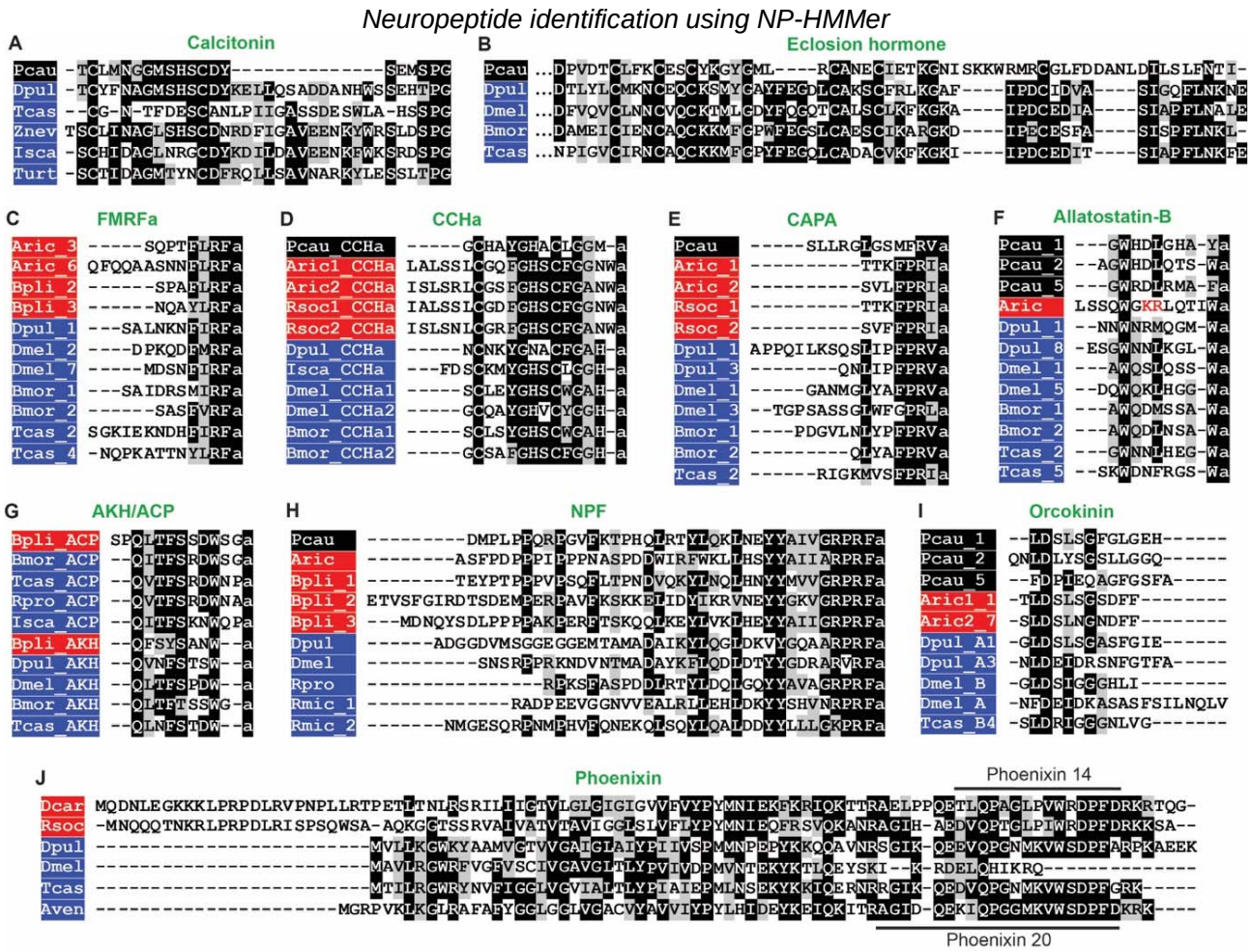

Figure 3: Multiple sequence alignments of neuropeptides discovered in Priapulida and Rotifera. Alignments of **(A)** calcitonin, **(B)** eclosion hormone, **(C)** FMRFamide (FMRFa), **(D)** CCHamide (CCHa), **(E)** CAPA, **(F)** allatostatin-B, **(G)** adipokinetic hormone (AKH) and AKH/corazonin-related peptide (ACP), **(H)** neuropeptide F (NPF), **(I)** orcokinin and **(J)** phoenixin neuropeptides. Only mature peptides are shown except for eclosion hormone (partial alignment) and phoenixin (complete neuropeptide precursor). A dibasic cleavage site (KR) in *Adineta ricciae* allatostatin-B is shown in red. These residues are likely cleaved to generate two peptides (LSSQWamide and LQTIWamide). Priapulid species are colored in black, rotifers in red, and arthropods in blue. Conserved residues are highlighted in black or gray. Species names: Pcau, *Priapulis caudatus*; Dcar, *Didymodactylos carnosus*; Rsoc, *Rotaria socialis*; Aric, *Adineta ricciae*; Bpli, *Brachionus plicatilis*; Dpul, *Daphnia pulex*; Bmor, *Bombyx mori*; Tcas, *Tribolium castaneum*; Rpro, *Rhodnius prolixus*; Dmel, *Drosophila melanogaster*; Isca, *Ixodes scapularis*; Rmic, *Rhipicephalus microplus*; Znev, *Zootermopsis nevadensis*; Turt, *Tetranychus urticae*; Aven, *Araneus_ventricosus.*

Table 1: Summary of neuropeptides identified in Priapulida and Rotifera

	Priapulida	Rotifera			
Precursors / Species	***Priapulus caudatus***	***Adineta ricciae***	***Brachionus plicatilis***	***Didymodactylos carnosus***	***Rotaria socialis***
ACP/APWGa			1		
AKH			1		
Agatoxin	2				
Allatostatin-A		1	1	2	2
Allatostatin-B / Wamide	1	1		1	
Allatostatin-C/CC/CCC	2	1			
Bursicon alpha	2	1			
Bursicon beta	1				
Calcitonin	1				
CAPA	1	1			1
CCAP	1			2	1
CCHa	1	2			2
CNMa	1				
DH31	1				
DH44	1				
ETH	1				
Eclosion Hormone	3				
FMRFa		1	1	2	2
GPA2	2				2
GPB5	1				
Insulin/IGF-like	2	1			1
ITP	1				
Leucokinin				2	1
Neuropeptide-F	1	5	3	1	3
NUCB	1	2	1	1	2
Orcokinin-A/B	1	2		1	2
PDF		1			
Phoenixin				1	1
Proctolin		1			
PTTH	1				
RYa/Luqin	1				
Tachykinin	1				
TRH	1				
Vasopressin/Oxytocin		2	1		2

Table S1: Summary of neuropeptides recovered using NP-HMMer in proteomes of model arthropods. Closely-related neuropeptides identified for a given model are included in brackets. Neuropeptides highlighted in yellow indicate sequences that were not part of the proteome used for analysis but were successfully recovered by NP-HMMer following their addition to the proteome. Phoenixin neuropeptide was identified for the first time in *Drosophila melanogaster*.

Precursors / Species	*Drosophila melanogaster*	*Daphnia pulex*	*Tribolium castaneum*
ACP	1 (1 AKH)	1 (1 AKH)	3 (2 AKH)
AKH	1	1	3 (1 ACP)
Agatoxin	0	1	1
Allatostatin-A	1	1	0
Allatostatin-B	1	1	2 (1 Tachykinin)
Allatostatin-C/CC/CCC	2 (1 Allatostatin-CC)	1 (1 Allatostatin-CCC)	2 (1 Allatostatin-CC)
Allatotropin	0	1	1
Bursicon alpha	1	2 (1 Bursicon beta)	2 (1 Bursicon beta)
Bursicon beta	1	2 (1 Bursicon alpha)	2 (1 Bursicon alpha)
Calcitonin	0	1	1
CAPA	1	1	2 (1 PBAN)
CCAP	1	1	1
CCHa-1	2 (1 CCHa-2)	1 (1 CCHa)	2 (1 CCHa-2)
CCHa-2	2 (1 CCHa-1)	1 (1 CCHa)	2 (1 CCHa-1)
CNMa	1	1	1
Corazonin	1	1	1 (1 ACP)
DH31	1	1	1
DH44	1	1	2
ETH	1	1	1
Eclosion Hormone	1	2	2
Elevenin	0	1	1
FMRFa	1	1	1
GPA2	1	1	1
GPB5	1	1	1
Insulin/IGF-like	6	3	4
ITP	1	3	1
Leucokinin	1	0	0
Myosuppressin	1	1	1
Natalisin	1	1	2 (1 Tachykinin)
Neuroparsin	0	1	1
Neuropeptide-F	1	1	0
NUCB	1	1	1
Orcokinin-A/B	1 (1 Orcokinin-A)	1 (1 Orcokinin-A)	1 (1 Orcokinin-B)
PDF	1	1	1
Phoenixin	1	1	1
Proctolin	1	1	1
PTTH	1	0	1
RYa/Luqin	1	1	1
SIFa	1	1	1
sNPF	1	1	1
Sulfakinin	1	1	1
Tachykinin	1	2 (1 Natalisin)	2 (1 Allatostatin-B)
TRH	0	1	0
Trissin	1	0	1
Vasopressin/Oxytocin	0	1	1