Published in final edited form as:

J Proteome Res. 2011 July 1; 10(7): 3098–3106. doi:10.1021/pr2001176.

Discovery of Neuropeptides in the Nematode *Ascaris suum* by Database Mining and Tandem Mass Spectrometry

Jessica L. Jarecki¹, Brian L. Frey², Lloyd M. Smith², and Antony O. Stretton^{1,3,*}

¹Neuroscience Training Program, University of Wisconsin-Madison

²Department of Chemistry, University of Wisconsin-Madison

³Department of Zoology, University of Wisconsin-Madison

Abstract

Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) was used to discover peptides in extracts of the large parasitic nematode *Ascaris suum*. This required the assembly of a new database of known and predicted peptides. In addition to those already sequenced, peptides were either previously predicted to be processed from precursor proteins identified in an *A. suum* library of expressed sequence tags (ESTs), or newly predicted from a library of *A. suum* genome survey sequences (GSSs). The predicted MS/MS fragmentation patterns of this collection of real and putative peptides were compared with the actual fragmentation patterns found in the MS/MS spectra of peptides fractionated by MS; this enabled individual peptides to be sequenced. Many previously identified peptides were found, and 21 novel peptides were discovered. Thus, this approach is very useful, despite the fact that the available GSS database is still preliminary, having only 1X coverage.

Keywords

Neuropeptide; liquid chromatography-tandem mass spectrometry; nematode; *Ascaris suum*; genomic survey sequences; peptide prediction

Introduction

Across animal phyla, neuropeptides have been shown to have strong effects on a wide variety of behaviors¹. In order to understand the cellular mechanisms of such behaviors, a detailed knowledge of the circuitry is required. This includes a description of the neuromodulators present in the system, and of their effects on the neurons and synapses in each circuit. The parasitic nematode *Ascaris suum*, with only 298 neurons, is a simple system that allows for such detailed analysis. Extensive physiological characterizations of motoneurons and muscles have been performed, as have descriptions of classical neurotransmitters such as acetylcholine, gamma-amino butyric acid (GABA), dopamine, and serotonin²⁻⁸. It is also clear that there is a large array of neuropeptides present. Already, forty members of the *Ascaris* FMRFamide-related family of peptides (AF peptides, typically containing an -RFamide C-terminus) have been isolated, and have been shown to have potent effects on body posture, muscle tone, neuronal physiology, or second messenger systems⁹⁻²⁰. However, it is clear from mass spectrometry (MS)²¹⁻²³, and from immunochemical studies with anti-peptide antibodies²⁴, that there are still many peptides to

*Correspondence: Antony Stretton Department of Zoology 1117 W Johnson St. Madison WI, 53706. Tel: 608-262-2172; Fax: 608-262-9083 aostrett@wisc.edu.

be discovered – we estimate that there are at least 250 neuropeptides in *A. suum*²³, and a comparable number has been predicted in the small free-living nematode *Caenorhabditis elegans*²⁵.

There are several methods available for neuropeptide discovery. Some are direct, such as the Edman degradation of pure peptides from chemical fractionation of low molecular weight extracts of tissues, or the mass spectrometric analysis of nervous tissue or single neurons, both of which have yielded the sequences of peptides that are expressed *in vivo*^{12, 13, 20-22, 26}. Indirect methods involve prediction of peptides that are cleaved from precursor protein sequences deduced from cDNA (including expressed sequence tags – ESTs^{27, 28}) or from genomic sequences^{29, 30}. The validity of these predictions depends on several factors. The first is the ability to recognize neuropeptide precursor genes. Typically these genes encode proteins with an N-terminal signal peptide sequence that delivers them to the secretory pathway, and with predictable cleavage sites for proteolytic processing into shorter peptides (usually basic amino acids, in pairs or singly, flanking each processed peptide)²⁵. The second factor is the validity of the predictions about sites of cleavage, which depends on assumptions about the nature and cellular expression patterns of the cleavage enzymes. Unfortunately, these enzymes exist in multiple forms with differing specificity, and their cellular expression patterns are usually unknown. This problem is compounded by any covalent post-translational modifications of the original sequences, and by the existence of unusual cleavages that are exceptions to the basic amino acid rule³¹. Consequently, the indirect prediction of peptide sequences must remain tentative until proven by one or more of the direct methods.

In principle, recent advances in liquid chromatography-tandem mass spectrometry (LC-MS/MS) should make the description of the peptidome of any organism relatively easy. This approach has been applied very successfully for protein identification in the field of proteomics by enzymatically digesting the complex mixtures of proteins into peptides prior to LC-MS/MS analysis³²⁻³⁴. The pattern of peptide fragment peaks in the MS/MS spectra are matched to theoretical peptide spectra, which have been predicted from a protein database derived from a sequenced genome, annotated so that the possible protein products have been identified. This process of database searching relies heavily upon knowledge of the cleavage characteristics of the enzyme used to digest the proteins into peptides - trypsin is the most common, but other enzymes are suitable as well. Assigning a peptide sequence to an experimental spectrum by using this database searching algorithm is much easier than interpreting the MS/MS spectra with no previous sequence information, a process termed *de novo* sequencing.

Applying this technique of database searching to the discovery of neuropeptides in *A. suum* is not straightforward, however. Genomic sequences have only recently become available for *A. suum*, and they consist of genome survey sequences (GSSs), with only 1 X coverage (M. Mitreva, personal communication). The sequences are not assembled, and sequencing errors have not been corrected. Nevertheless, GSSs have been extremely useful in aiding *de novo* sequencing, and we have used them in conjunction with MS/MS data to sequence many novel peptides from single neurons^{21, 35}. In addition, neuropeptides are not products of tryptic cleavage, so the algorithms typically used to simplify large database searches cannot be used. Instead, we assembled a smaller database to improve peptide identification. This new database consisted of translated GSS or cDNA sequences that encoded predicted or known peptides. These sequences were either known transcripts from *A. suum*^{20, 21, 23, 28, 35-37, 44}, were identified as sequelogs of known or predicted *C. elegans* peptide-encoding ESTs^{27, 28}, or were discovered from BLAST searches of *A. suum* GSSs. The goal of the current study was to explore whether the predicted peptides are expressed at detectable levels in *A. suum*.

This general approach has already been used very successfully in exploring the neuropeptidomes of *Drosophila melanogaster*³⁸ and *C. elegans*³⁹, taking advantage of completely sequenced genomes, and extensive predictions of putative peptides^{29, 30}; in *C. elegans* the number of sequenced peptides was increased from 10 to 64³⁹. Relatively, we are at a disadvantage in *A. suum*, because of the restricted nature of the available nucleic acid sequence data, and we wished to determine whether these data were sufficient to give useful information (see also ref. 40). The initial use of the new database with LC-MS/MS data, reported here, has discovered 21 novel peptides in *A. suum*, a substantial addition to the previous total of 50 peptides sequenced in the last two decades^{12, 13, 20-23, 26, 35}. In addition to the new peptides, many previously identified peptides were also found. Because it is relatively rapid, this method is indeed a useful complement to other methods of peptide discovery. Among the novel sequenced peptides are 15 peptides that are not members of the AF peptide family, since they lack the common C-terminal sequence (-RFamide) that defines the AF peptides. This makes the further exploration of the nature of this class of non-AF peptides in *A. suum*, and an analysis of their functional roles, an important priority.

Materials and Methods

Animals

A. suum were harvested from the small intestines of pigs at a regional slaughterhouse. The worms were maintained for up to 3 days at 37 °C in phosphate-buffered saline (PBS; 8.5 mM sodium phosphate, 150 mM sodium chloride, pH 7.4). The head region (anterior 1-2 cm) was removed and immediately frozen on dry ice. The heads were stored at -80 °C until use.

Peptide Extraction and Sample Preparation

Heads (N = 2,500) were freeze-powdered, and then peptides extracted with 500 mL ice-cold methanol: water: acetic acid, 90:9:1. The extract was centrifuged to remove solids, and the supernatant was then concentrated by rotary evaporation. The concentrated extract (4 mL) was diluted with 500 mL 0.1% aqueous trifluoroacetic acid (TFA). Lipids were removed by extraction with 4 × 100 mL 15:85 ethyl acetate:hexanes. Aliquots (60 mL) of the remaining aqueous solution were applied to 8 activated C18 cartridges (360 mg, SepPak, Waters Associates, Inc., Milford, MA). After washing each cartridge with 0.1% aqueous TFA, peptides were eluted with 5 mL 1:1 acetonitrile (ACN): water containing 0.1% TFA. The fractions were concentrated to ca. 500 µL in a vacuum centrifuge, and a Millipore YM3 Size Exclusion column was used to remove proteins. The resultant "low molecular weight fraction" was dried almost to completion and then diluted in 50 µL 0.1% formic acid in 1% aqueous ACN. However, proteins were still prevalent so additional size-exclusion chromatography was performed. A 25 µL aliquot was separated on a 200 × 4.6 mm PolyHYDROXYETHYL A column (PolyLC, Columbia, MD) using a high performance liquid chromatograph (HPLC; Beckman, Fullerton, CA) controlled by System Gold software. The mobile phase was 50 mM aqueous ammonium acetate, with a flow rate of 0.1 mL/min. All fractions eluting prior to 25 min. were collected by hand, combined, dried almost to completeness, and resuspended in 20 µL 0.1% formic acid in 1% aqueous ACN.

LC- MS/MS Analysis

The neuropeptide sample was diluted 3-fold with 0.1% formic acid:2% aqueous ACN, and 3 µL injected into the LC-MS/MS capillary system consisting of an HPLC (Waters nanoAcquity, Milford, MA) connected to an electrospray ionization FT/ion-trap mass spectrometer (LTQ orbitrap Velos, Thermo Fisher Scientific, San Jose, CA). The sample loading occurred for 20 min at a flow-rate of 1 µL/min onto a trapping column prepared by packing 5 cm of 5 µm-diameter 300 Å-pore C18 beads (Western Analytical Products, Inc.

Murrieta, CA) into a $75 \times 365 \mu\text{m}$ fused silica capillary having a frit at one end. The analytical column ($50 \times 365 \mu\text{m}$) was packed with 15 cm of the same C18 beads, but instead of packing against a frit, the capillary tip was pulled to $\sim 1 \mu\text{m}$ with a P-2000 pipette puller (Sutter Instruments Co.). The peptides were eluted at a flow rate of $0.2 \mu\text{L}/\text{min}$ in an aqueous mobile phase containing 0.1% formic acid with a gradient of increasing ACN: 2% to 15% in the first 15 min. and then up to 45% in the next 75 min (Run 1), 90 min (Run 2), or 105 min (Runs 3 and 4).

A survey mass scan ($150\text{--}2000 m/z$ – mass to charge ratio) was performed in the FT orbitrap at a resolution of 60,000. Various types of MS/MS scans were performed: collision induced dissociation (CID) at 30% relative collision energy, high energy collision dissociation (HCD) at 35 or 45% relative collision energy, and electron transfer dissociation (ETD) for 70 ms. These dissociation methods are somewhat complementary in that different methods lead to optimal fragmentation for various types of peptides depending upon their mass, ion charge, and amino acid sequence. The isolation width for the parent ions was $3.0 m/z$, and the product ions from CID, HCD, or ETD fragmentation were analyzed in the FT orbitrap detector at a resolution of 7500.

The MS/MS conditions for the four runs were as follows (percentages refer to relative collision energy):

Run 1: CID at 30% and HCD at 45% on all +2 and higher charge ions. The 5 most intense peaks from the MS survey scan were chosen for fragmentation by both CID and HCD.

Run 2: CID at 30%, HCD at 35%, HCD at 45%, and ETD 70 ms on all assigned charge state ions +1 and higher. The 4 most intense peaks were chosen for fragmentation by all 4 of these MS/MS types.

Run3: CID at 30% and HCD at 35% on +1 and +2 ions. The 6 most intense peaks were chosen for both MS/MS types.

Run 4: HCD at 45% and ETD 70 ms on +3 and higher charge ions. The 6 most intense peaks were chosen for both MS/MS types.

Database creation

Translated Basic Local Alignment Search Tool (tBLASTn) searches of *A. suum* GSSs were performed in September 2010 using the National Center for Biotechnology Information (NCBI) BLAST server (<http://www.ncbi.nlm.nih.gov/BLAST>) as described previously by McVeigh et al., (2008) ²⁷. Lists of *C. elegans* *nlps* ^{25, 27, 30} were used to develop search strings. Genes encoding multiple peptides were concatenated into a single string. The Expect value was set to 30,000, with a word size of 2, no compositional adjustments, and complexity filter off. Results were translated (<http://www.expasy.ch/tools/dna.html>) and examined by eye for neuropeptide motifs and cleavage sites. Previously cloned and/or predicted *afps* (*Ascaris* FMRFamide-like precursor proteins) and *As-nlps* (*Ascaris* neuropeptide-like precursor proteins) were also added to the database ^{20, 21, 23, 27, 28, 35-37, 44}.

Peptide identification

The MS/MS data of all 4 runs were searched against the *A. suum* peptide database using SEQUEST or Mascot on in-house servers. These bioinformatics programs match the fragmentation data against any FASTA format protein database ^{40, 41}. To use Mascot, data files were transformed to mzXML using ReAdW and then to mgf using TPP (<http://tools.proteomecenter.org/software.php>). Multiple MS/MS ion searches were

performed using no C-terminal modifications or a fixed C-terminal amidation in combination with variable modifications of pyroglutamic acid (Glu), acetylation (Lys or N-terminus), and oxidation (Met), which are all common post-translational modifications of neuropeptides. Tentative peptide identifications, which had Expect values <0.5 and a significance threshold of $p < 0.05$ (Mascot) or XCorr values >2 (SEQUEST) and a parent mass within 5 ppm of its expected value, were further explored manually. To confirm these identifications as naturally occurring peptides, stringent requirements had to be met: 1) the peptide is flanked in the precursor by basic cleavage sites or N-terminally flanked by the signal peptide, and 2) if the peptide sequence contains a C-terminal amidation, the precursor peptide sequence includes a C-terminal glycine.

RESULTS AND DISCUSSION

This study attempts to identify novel neuropeptides in *A. suum* through BLAST (Basic Local Alignment Search Tool) searches of genomic survey sequences (GSSs) coupled with MS/MS analysis of a peptide extract.

GSS Database search

Previously, mining of the nematode expressed sequence tag (EST) databases has allowed the cross-species identification of a large number of predicted -RFamide peptides, as well as members of the neuropeptide-like protein (NLP) family (peptides that do not contain a C-terminal -RFamide) including several in *A. suum*²⁷. In this study, the *A. suum* GSS database was mined using *C. elegans* peptide-encoding transcripts to discover novel *A. suum* putative peptides. This was done both to identify possible novel peptides and to generate a reference database for MS analysis. Known translated *A. suum* peptide-encoding transcripts and ESTs were also added to this database. When analyzing GSSs as compared to ESTs, there is an added complication since ESTs are derived from the fully processed mRNA and GSSs are derived from genomic DNA. The GSSs contain introns and are not assembled, unlike completely sequenced genomes, making prediction of the final peptide far more difficult. With this in mind, more BLAST search results were included in the database used for MS analysis than those that included predicted peptides. Despite these complications, we predicted 12 novel putative - RFamide peptides from 9 GSSs (Table 1). Seven of these GSSs are sequelogs of known *C. elegans flps* (FMRFamide-like peptide transcripts). Two peptides, predicted from a search using Ce-FLP-10 as the query, have no obvious sequelog in *C. elegans*. We also found 44 *nlp*s predicted from 21 GSSs (Tables 2 and 3). This number does not include sequelogs of the anti-microbial *nlp*s, *Ce-nlp-24* - *Ce-nlp-33*, because it is difficult to predict peptide cleavage in these sequences. However, the 32 related GSSs that encode these antimicrobial *nlp*s were included in the database used for MS analysis. Several searches returned more than one GSS encoding a peptide related to the peptide query. For example, the search using Ce-NLP-15 as the query returned two GSSs, *As-nlp-15a* and *As-nlp-15b*, each encoding the peptide AFDSLGTGAGFTGFD and two different additional sequence-related peptides. It is unclear if the transcripts from these two GSSs are spliced together to form one transcript or if this is a gene duplication. Similarly, searches with Ce-NLP-18 and Ce-NLP-20 each returned two distinct peptide-encoding GSSs, but the high degree of similarity between first the *C. elegans* peptides and then the newly predicted *A. suum* peptides can cause difficulty in naming these peptides and peptide-encoding sequences by similarity, as has become the convention. Once the transcripts are cloned, more permanent names for transcripts and peptides will be assigned.

Mass spectrometric analysis of peptide extracts

Peptides extracted from freeze-powdered *A. suum* heads were analyzed in four runs by LC-MS/MS (liquid chromatography-tandem mass spectrometry). Each run produced a complex

base-peak chromatogram containing numerous overlapping peaks. Survey mass spectra were automatically acquired to find suitable targets for MS/MS, and the appropriate m/z ions were selected for fragmentation (Figure 1). The fragmentation patterns in the MS/MS spectra were then compared, by Mascot or SEQUEST search engines, to predicted fragmentation patterns from the database described above. The selection and fragmentation conditions were varied in the four runs in an attempt to maximize the number of peptide identifications (See Methods for details).

AF peptide identification

In this study, members of 6 AF peptide families were identified although we did not identify any of the newly predicted AF peptides (Table 4). In all but one of these families, more peptides are known to exist than were identified here; these peptides may have been lost during sample preparation or due to insufficient chromatographic separation. The sequences of several peptides were confirmed with our results, although they were previously predicted and analyzed by MS (not MS/MS)²³. These include a peptide resulting from an incomplete cleavage of AF3, ENEKKAVPGVLRFa (AF3*), and several members of the *afp-5* (*As-nlp-13*) family of peptides: SNAFDRNFmNFa, SDAFSRNfNFa, and ESQFSRDFLNfFa (The “m” refers to an oxidized methionine). Although these peptides are not actually RFamides, we have included them as AF peptides here because another peptide encoded by this transcript, AF17, which has an HFamide rather than an RFamide C-terminus, was originally identified by its RFamide immunoreactivity, and therefore named an AF peptide⁴³. Peaks with m/z values corresponding to the predicted mass of all of these peptides in an unoxidized form have been seen in MALDI-MS of whole *A. suum* ganglia and single neurons^{21, 23}. We believe that the oxidation is an artifact of the isolation procedure, since no effort was made to keep the peptides reduced during the preparative steps prior to LC-MS/MS.

A peptide with sequence LmDPLIRFa was identified, which is interesting because it is a fragment of peptide AF34 (DSKLMDPLIRFa) encoded by the *afp-13* transcript. The AFP-13 precursor protein, which encodes several sequenced peptides, is known to have both single and double basic amino acid cleavage sites²¹, so it is plausible that AF34 undergoes a second cleavage at the internal K. However, MS of the isolated ALA neuron, which contains AF34 as well as the other members of the *afp-13* peptide family, did not reveal a peak corresponding to the shorter peptide²¹. This could be a case of cell-specific cleavage of these peptides; alternatively, this peptide could arise from degradation during storage, extraction, or fractionation preceding MS analysis. We have had previous indications that proteolytic degradation may occur: AF10 (GFGDEMSMPGVLRFa) was originally discovered by Edman degradation of highly purified peptides²⁶; in the same purification, a related peptide, named AF12 (FGDEMSMPGVLRFa), was also isolated. However, when the *afp-1* transcript was subsequently cloned and sequenced, there was no separately encoded AF12 sequence - the AF12 sequence occurs only within AF10⁴⁴. Furthermore, AF12 was not detected in MS of whole ganglia²³, so we believe that AF12 is a degradation product of AF10, probably due to the action of an aminopeptidase. Interestingly, in addition to AF12, we have now found two more truncated forms of AF10 (Table 4). In earlier fractionations of peptide extracts, truncated forms of other peptides derived from *afp-1* were detected (with single amino N-terminal truncations of peptides AF3, AF13, AF14, and AF20 – Cowden and Stretton, unpublished). It is striking that, so far, N-terminal loss is particularly prevalent among the products of the *afp-1* transcript; it is unclear whether these changes occur in the cells expressing these peptides, or whether they represent unusual susceptibility to the action of extracellular aminopeptidases, which have been shown to be present in *A. suum*⁴⁵. One other case of N-terminal truncation has been observed, the loss of the N-terminal D from DFDRDFMHFa to generate AF17 (FDRDFMHFa). However, in this

case there is strong evidence that this is due to an unusual susceptibility of the N-terminal D to loss under the acidic conditions of the extraction ³⁶.

As-NLP peptide identification

A total of five As-NLP peptides (Ascaris neuropeptide-like peptides), encoded by sequences found in the EST library, were identified (Table 4). One of these arises from *As-nlp-7* (NHLIGFDDPRLFSSSYa, Figure 2a) and four are encoded by *As-nlp-3*, two of which were predicted by McVeigh et al., (2008) ²⁷. Of the two not predicted, As-NLP-3-4* is the result of an incomplete peptide cleavage. The other peptide appears to be the result of an atypical cleavage after an alanine, which we believe might be due to cleavage of the signal peptide. The flaw in this hypothesis is that the EST does not include an initiating methionine, but if one were added at the beginning of the translated gene and analyzed by SignalP ⁴⁶, the signal peptide is predicted to be cleaved to form the N-terminus of As-NLP-3-1. We are currently attempting to clone the full-length transcript to resolve this issue.

Peptides encoded by 6 distinct GSSs were also identified (Table 4). Nine peptides were previously described from the *As-nlp-8/anp-1* GSS ³⁵, and three of those were found in this study along with an additional peptide fragment. This fragment is likely the result of a procedural artifact, as was seen with several AF peptides. Similar results were observed for the highly related peptides from *As-nlp-15*: one full length peptide (SLDSPANRGFTHFD, Figure 2b), and two truncated peptides missing two C-terminal amino acids in this case. From *As-nlp-21*, one of three predicted peptides was identified. From the GSS *As-nlp-5*, two peptides were detected (TSISRSKDAKKPSLDILEGAGFSPL and ALDDLEGVGFGGmL), although more were predicted. The former is possibly the result of a signal peptide cleavage and a missed internal cleavage, and the latter is a more conventional peptide. Two peptides were also identified from *As-nlp-9* and *As-nlp-23*, which were not the products that would have originally been predicted, although they could be products of conventional cleavages. One of these (FGSFSPYYFYQQPK, Figure 2c) would more likely have been processed to remove the C-terminal lysine, according to conventional wisdom ²⁵. However, it is possible that the proline before the lysine interfered with this “conventional” processing by carboxypeptidase. For the other peptide, SLALGRVDFRPG (Figure 2d), the typical prediction of the final peptide product would be SLALGRVDFRPa. This may result from proline interference with the amidating enzymes; alternatively, it is possible that both forms exist but we did not detect the amidated form. In our recent single cell MALDI-MS experiments, the non-amidated C-terminal glycine form of several peptides have been detected, and they are presumed to be intermediates in the processing cascade ²¹. Interestingly, in experiments with *C. elegans* peptide extracts, no Pamide peptides have been detected, although many are predicted ²⁵.

Prediction versus chemical sequencing

As has been shown with several of the peptides identified in this study, peptide prediction can be helpful but also occasionally inaccurate. We found peptide prediction essential for the type of MS analysis performed here because the very large number of spectra generated made manual interpretation infeasible—Mascot and/or SEQUEST searches were required. However, the extensive size of a total GSS or EST database, when used in these searches, either caused the program to crash or returned very few identifications. Thus, having removed non-coding genomic sequences and sequences not identified as putative peptides, the smaller database containing only predicted peptides of interest allowed a more practical and confident route to the identification of peptides. An analogy may be to consider the greater ease of searching for peptide “needles” within a sequence “haystack” when much of the non-interesting “hay” has been removed.

Are there more peptides?

Although many novel peptides were discovered in this study, it is clear that there are many more peptides that remain to be sequenced. The mass spectra in these experiments show complex mixtures of ions entering the instrument at one time; thus, better chromatographic separation prior to MS analysis would help identify a larger number of peptides. It is striking that AF2 was not detected in this study, even though it is believed to be the most abundant peptide in *A. suum* and it has been detected repeatedly in whole ganglia and single cell MALDI-MS^{21, 23, 35}. Interestingly, no peptides without basic amino acids have been detected by MALDI-MS, but three were detected in this study.

CONCLUSIONS

The use of GSS database mining in combination with MS yielded a dramatic increase in the number of predicted and identified peptides in *A. suum*. This was especially true of the *As-nlps*, of which only the *As-nlp-8 (anp-1)* family of peptides had been sequenced previously³⁵. The prediction and identification of all of these peptides is an important first step in understanding the vital role neuropeptides play in the nervous system of *A. suum*.

Acknowledgments

We thank Makedonka Mitreva at Washington University, St Louis MO, for access to her large GSS database, Bill Feeny for help with the figures, Lakshmi Narayanan Bairavasundaram for his computer programming assistance, and Greg Sabat and the University of Wisconsin Biotechnology Center for assistance with, and access to, Mascot. We also thank Philippa Claude for critically reviewing the manuscript. This research was supported by National Institutes of Health Grants RO1 AI15429 to AOS, and P01 GM081629 to LMS. JLJ was supported by training grant NRSA T32 GM007507.

ABBREVIATIONS

ACN	acetonitrile
AF	<i>Ascaris</i> FMRFamide-related
afp	<i>Ascaris</i> FMRFamide-like precursor protein
BLAST	Basic Local Alignment Search Tool
CID	collision induced dissociation
EST	expressed sequence tag
ETD	electron transfer dissociation
flp	FMRFamide-like peptide
FT	Fourier transform
GABA	gamma-amino butyric acid
GSS	genome survey sequence
HCD	high energy collision dissociation
HPLC	high performance liquid chromatography
LC-MS/MS	liquid chromatography coupled with tandem mass spectrometry
m/z	mass to charge ratio
NLP	neuropeptide-like protein
TFA	trifluoroacetic acid

REFERENCES

1. Strand, FL. Neuropeptides. MIT Press; Cambridge, MA: 1999.
2. Del Castillo J, De Mello WC, Morales T. The initiation of action potentials in the somatic musculature of *Ascaris lumbricoides*. *J Exp Biol.* 1967; 46:263–279. [PubMed: 6033994]
3. Davis RE, Stretton AO. Passive membrane properties of motoneurons and their role in long-distance signaling in the nematode *Ascaris*. *J Neurosci.* 1989; 9:403–414. [PubMed: 2918369]
4. Davis RE, Stretton AO. Signaling properties of *Ascaris* motoneurons: graded active responses, graded synaptic transmission, and tonic transmitter release. *J Neurosci.* 1989; 9:415–425. [PubMed: 2563763]
5. Johnson CD, Stretton AO. GABA-immunoreactivity in inhibitory motor neurons of the nematode *Ascaris*. *J Neurosci.* 1987; 7:223–235. [PubMed: 3543249]
6. Johnson CD, Stretton AO. Localization of choline acetyltransferase within identified motoneurons of the nematode *Ascaris*. *J Neurosci.* 1985; 5:1984–1992. [PubMed: 4020429]
7. Johnson CD, Reintz CA, Sithigorngul P, Stretton AO. Neuronal localization of serotonin in the nematode *Ascaris suum*. *J Comp Neurol.* 1996; 367:352–360. [PubMed: 8698897]
8. Sulston J, Dew M, Brenner S. Dopaminergic neurons in the nematode *Caenorhabditis elegans*. *The Journal of comparative neurology.* 1975; 163:215–226. [PubMed: 240872]
9. Bowman JW, Friedman AR, Thompson DP, Ichhpurani AK, Kellman MF, Marks N, Maule AG, Geary TG. Structure-activity relationships of KNEFIRFamide (AF1), a nematode FMRFamide-related peptide, on *Ascaris suum* muscle. *Peptides.* 1996; 17:381–387. [PubMed: 8735963]
10. Brownlee D, Holden-Dye L, Walker R. The range and biological activity of FMRFamide-related peptides and classical neurotransmitters in nematodes. *Adv Parasitol.* 2000; 45:109–180. [PubMed: 10751940]
11. Brownlee DJ, Fairweather I, Holden-Dye L, Walker RJ. Nematode neuropeptides: Localization, isolation and functions. *Parasitol Today.* 1996; 12:343–351. [PubMed: 15275172]
12. Cowden C, Stretton AO. AF2, an *Ascaris* neuropeptide: isolation, sequence, and bioactivity. *Peptides.* 1993; 14:423–430. [PubMed: 8332542]
13. Cowden C, Stretton AO, Davis RE. AF1, a sequenced bioactive neuropeptide isolated from the nematode *Ascaris suum*. *Neuron.* 1989; 2:1465–1473. [PubMed: 2627377]
14. Davis RE, Stretton AO. Structure-activity relationships of 18 endogenous neuropeptides on the motor nervous system of the nematode *Ascaris suum*. *Peptides.* 2001; 22:7–23. [PubMed: 11179593]
15. Fellowes RA, Maule AG, Marks NJ, Geary TG, Thompson DP, Shaw C, Halton DW. Modulation of the motility of the vagina vera of *Ascaris suum* in vitro by FMRF amide-related peptides. *Parasitology.* 1998; 116(Pt 3):277–287. [PubMed: 9550221]
16. Holden-Dye L, Brownlee DJ, Walker RJ. The effects of the peptide KPNFIRFamide (PF4) on the somatic muscle cells of the parasitic nematode *Ascaris suum*. *Br J Pharmacol.* 1997; 120:379–386. [PubMed: 9031739]
17. Maule AG, Geary TG, Bowman JW, Marks NJ, Blair KL, Halton DW, Shaw C, Thompson DP. Inhibitory effects of nematode FMRFamide-related peptides (FaRPs) on muscle strips from *Ascaris suum*. *Invert Neurosci.* 1995; 1:255–265. [PubMed: 9372147]
18. Maule AG, Shaw C, Bowman JW, Halton DW, Thompson DP, Geary TG, Thim L. KSAYMRFamide: a novel FMRFamide-related heptapeptide from the free-living nematode, *Panagrellus redivivus*, which is myoactive in the parasitic nematode, *Ascaris suum*. *Biochem Biophys Res Commun.* 1994; 200:973–980. [PubMed: 8179635]
19. Reintz CA, Herfel HG, Messinger LA, Stretton AO. Changes in locomotory behavior and cAMP produced in *Ascaris suum* by neuropeptides from *Ascaris suum* or *Caenorhabditis elegans*. *Mol Biochem Parasitol.* 2000; 111:185–197. [PubMed: 11087928]
20. Yew JY, Davis R, Dikler S, Nanda J, Reinders B, Stretton AO. Peptide products of the *afp-6* gene of the nematode *Ascaris suum* have different biological actions. *J Comp Neurol.* 2007; 502:872–882. [PubMed: 17436302]
21. Jarecki JL, Andersen K, Konop CJ, Knickelbine JJ, Vestling MM, Stretton AO. Mapping neuropeptide expression by mass spectrometry in single dissected identified neurons from the

- dorsal ganglion of the nematode *Ascaris suum*. *ACS Chem Neurosci*. 2010; 1:505–519. [PubMed: 20806053]
22. Yew JY, Dikler S, Stretton AO. De novo sequencing of novel neuropeptides directly from *Ascaris suum* tissue using matrix-assisted laser desorption/ionization time-of-flight/time-of-flight. *Rapid Commun Mass Spectrom*. 2003; 17:2693–2698. [PubMed: 14673815]
 23. Yew JY, Kutz KK, Dikler S, Messinger L, Li L, Stretton AO. Mass spectrometric map of neuropeptide expression in *Ascaris suum*. *J Comp Neurol*. 2005; 488:396–413. [PubMed: 15973679]
 24. Sithigorngul P, Stretton AO, Cowden C. Neuropeptide diversity in *Ascaris*: an immunocytochemical study. *J Comp Neurol*. 1990; 294:362–376. [PubMed: 2341616]
 25. Li C, Kim K. Neuropeptides. *Wormbook*. 2008
 26. Cowden C, Stretton AO. Eight novel FMRFamide-like neuropeptides isolated from the nematode *Ascaris suum*. *Peptides*. 1995; 16:491–500. [PubMed: 7651904]
 27. McVeigh P, Alexander-Bowman S, Veal E, Mousley A, Marks NJ, Maule AG. Neuropeptide-like protein diversity in phylum Nematoda. *Int J Parasitol*. 2008; 38:1493–1503.
 28. McVeigh P, Leech S, Mair GR, Marks NJ, Geary TG, Maule AG. Analysis of FMRFamide-like peptide (FLP) diversity in phylum Nematoda. *Int J Parasitol*. 2005; 35:1043–1060. [PubMed: 16076468]
 29. Li C, Kim K, Nelson LS. FMRFamide-related neuropeptide gene family in *Caenorhabditis elegans*. *Brain Res*. 1999; 848:26–34. [PubMed: 10612695]
 30. Nathoo AN, Moeller RA, Westlund BA, Hart AC. Identification of neuropeptide-like protein gene families in *Caenorhabditis elegans* and other species. *Proc Natl Acad Sci U S A*. 2001; 98:14000–14005. [PubMed: 11717458]
 31. Fisher JM, Scheller RH. Prohormone processing and the secretory pathway. *J Biol Chem*. 1988; 263:16515–16518. [PubMed: 3053690]
 32. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003; 422:198–207. [PubMed: 12634793]
 33. Cravatt BF, Simon GM, Yates JR 3rd. The biological impact of mass-spectrometry-based proteomics. *Nature*. 2007; 450:991–1000. [PubMed: 18075578]
 34. Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science*. 2006; 312:212–217. [PubMed: 16614208]
 35. Jarecki, JL. PhD thesis. University of Wisconsin; Madison: 2010. Identification and Localization of Neuropeptides in *Ascaris suum* by Mass Spectrometry.
 36. Nanda, J. PhD thesis. University of Wisconsin; Madison: 2004. Molecular biological analysis of neuropeptide gene transcripts from the nematode, *Ascaris suum*.
 37. Nanda JC, Stretton AO. In situ hybridization of neuropeptide-encoding transcripts *afp-1*, *afp-3*, and *afp-4* in neurons of the nematode *Ascaris suum*. *J Comp Neurol*. 2010; 518:896–910. [PubMed: 20058230]
 38. Baggerman G, Boonen K, Verleyen P, De Loof A, Schoofs L. Peptidomic analysis of the larval *Drosophila melanogaster* central nervous system by two-dimensional capillary liquid chromatography quadrupole time-of-flight mass spectrometry. *J Mass Spectrom*. 2005; 40:250–260. [PubMed: 15706625]
 39. Husson SJ, Clynen E, Baggerman G, De Loof A, Schoofs L. Discovering neuropeptides in *Caenorhabditis elegans* by two dimensional liquid chromatography and mass spectrometry. *Biochem Biophys Res Commun*. 2005; 335:76–86. [PubMed: 16061202]
 40. Ma M, Bors EK, Dickinson ES, Kwiatkowski MA, Sousa GL, Henry RP, Smith CM, Towle DW, Christie AE, Li L. Characterization of the *Carcinus maenas* neuropeptidome by mass spectrometry and functional genomics. *Gen Comp Endocrinol*. 2009; 161:320–334. [PubMed: 19523386]
 41. Eng JK, McCormack AL, Yates JR. An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. *J Am Soc Mass Spectr*. 1994; 5:976–989.
 42. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 1999; 20:3551–3567. [PubMed: 10612281]

43. Davis RE, Stretton AO. The motornervous system of *Ascaris*: electrophysiology and anatomy of the neurons and their control by neuromodulators. *Parasitology*. 1996; 113(Suppl):S97–117. [PubMed: 9051930]
44. Edison AS, Messinger LA, Stretton AO. *afp-1*: a gene encoding multiple transcripts of a new class of FMRFamide-like neuropeptides in the nematode *Ascaris suum*. *Peptides*. 1997; 18:929–935. [PubMed: 9357048]
45. Sajid M, Isaac RE, Harrow ID. Purification and properties of a membrane aminopeptidase from *Ascaris suum* muscle that degrades neuropeptides AF1 and AF2. *Mol Biochem Parasitol*. 1997; 89:225–234. [PubMed: 9364967]
46. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*. 2004; 340:783–795. [PubMed: 15223320]

Figure 1.

Schematic overview of MS data acquisition. (a) Base peak chromatogram of Run 1. The red line indicates the time of the MS scan in b. (b) Example of MS survey scan in which the doubly charged ion 594.84 (AF19) was selected for fragmentation. (c) MS/MS spectrum of AF19 with the observed b and y fragment ions labeled. These fragment ions were matched to those predicted from the AF19 sequence.

Figure 2.

Four novel peptide MS/MS spectra with b and y fragment ions denoted. (a) CID spectrum of NHLIGDDPRLFSSSYa, m/z = 933.962, z = +2. (b) CID spectrum of SLDSPANRGFTTHFD, m/z = 782.363, z = +2. (c) CID spectrum of FGSFSPYYFYQPK, m/z = 879.912, z = +2. (d) HCD spectrum of SLALGRVDFRPG, m/z = 644.363, z = +2.

Table 1

Putative *A. suum flp* peptides predicted from tBLASTn searches of the Genbank GSS database. The numbers below each gene are accession numbers.

<u>gene</u>		<u>predicted sequence</u>	<u>cleavage site</u>
<i>As-flp-3</i>	KR	GGPLGTMRFG	KR
ED071134			
<i>As-flp-4</i>	RK	ASFIRFG	KR
ED359465	KR	KASFIRLG	–
<i>As-flp-10a</i>	RR	LPHMEKISGFIRFG	KR
ED453735			
<i>As-flp-10b</i>	R	SCYMIRFG	KR
ED395851			
<i>As-flp-10c</i>	KR	TVYAVIRFG	K
ED291226			
<i>As-flp-17</i>	KR	KSAFVRFG	R
ED065987	KR	KSSYIRFG	–
<i>As-flp-19</i>	KK	WASQLRLG	KR
ED386389	KR	ASWASKVRFG	–
<i>As-flp-25</i>	K	TSYQYFRFG	R
ED280015			
<i>As-flp-31</i>	KK	GRPRGPLRFG	KR
ED243419			

Table 2Putative *A. suum* *nlp-1* -- *nlp-17* peptides predicted from tBLASTn searches of the Genbank GSS database.

As-nlp-1	KR	MDASRFYLGFG	KR
ED075198			
<i>As-nlp-5</i>	A	TSISRSKDAKKPSLDILEGAGFSPL	KK
ED209100	KK	PSLDILEGAGFSPL	KK
	KR	ALDTLEGSGFGFS	KR
	KR	ALDDLEGVGFGGML	RK
	KR	ALDSLEGTDFGL	KK
	KR	ALDYLEGSGFGML	KR
	KR	AFDAIEGADIGFH	KR
	KR	ALDMLEGSGFGL	KK
	KR	ALDSLEGTGFGL	KK
<i>As-nlp-6</i>	KR	YASRGGFGQ	KR
ED364076	KR	ASMRFSRVAQA	K
<i>As-nlp-9</i>	KR	AGARSFGPRIYDVESYLYPTN	KR
ED378410	KR	FGSFSPYYFYQQPK	R
	KR	GGGRTFASYWIPPSGNG	R
<i>As-nlp-10</i>	KR	SSIPFHGGIYG	KR
ED196733	KR	TAALPFSGGLYG	KR
	KR	AGWIPFSGGLYG	KR
<i>As-nlp-15a</i>	KR	AFDSLGTAGFTGFD	KR
ED281083	KR	AFDSLGTSGFTGFD	KR
	KR	SLDSPANRGFTHFD	KR
<i>As-nlp-15b</i>	R	AFDSLGTAGFTGFD	KR
ED054924	KR	SFDSLNSRGFTGFD	KR
	KR	AFDSLVGHGFTGFD	-
<i>As-nlp-16</i>	R	NNDISSSGEHEHQ	KR
ED433984			
<i>As-nlp-17</i>	RK	NILSNMMRIG	RR
ED472655			

Table 3

Putative *A. suum* *nlp-18* - *nlp-23* and *nlp-34* - *nlp-46* peptides predicted from tBLASTn searches of the Genbank GSS database.

As-nlp-18a	RR	FAYSFAFA	KR
ED276235			
<i>As-nlp-18b</i>	KK	AGRFARHASGQQNELSFAFA	KR
ED229080	KR	APFA	KR
<i>As-nlp-20a</i>	KR	NKDANSVYAFSFA	KR
ED418549			
<i>As-nlp-20b</i>	K	AFANFAFA	KR
ED054056			
<i>As-nlp-21</i>	KR	KHYLQTDE	KR
ED417768	KR	GGGRSFRLSSLGE	KR
	KR	AGARSFPGAGGLT	R
<i>As-nlp-22</i>	KR	SIALGRFSLRPG	KR
ED456305			
<i>As-nlp-23a</i>	KR	SMALGRLAFRPG	KR
ED282494	KR	SLALGRVDFRPG	KR
	KR	NVDIADKDIEYLMEDSMG	KR
	KR	SAAFGRFHFRPG	KR
	KR	SLALGRSGFRPG	KR
	KR	SLALGRVGFRPG	KR
<i>As-nlp-23b</i>	K	GIAPYAGMRPG	KR
ED431569			
<i>As-nlp-35</i>	RR	DSKRAFLQQLMQRLKPRF	RR
ED247130			
<i>As-nlp-44</i>	RR	NDRPQSSAAVLVPYPRVG	KR
ED420546			
<i>As-nlp-46</i>	KR	NIAIGRGDGRPG	K
ED469442			

Table 4

A. suum peptides identified by either Mascot (M) or Sequest (S) searches in 4 different MS runs.

Gene	Peptide	Sequence	Run1		Run2		Run3		Run4	
			M	S	M	S	M	S	M	S
<i>afp-1</i>	AF3*	ENEKKA ⁺ PGVLR ⁺ Fa	+	+	+	+	+	+	+	+
	AF4	GDVPGVLR ⁺ Fa	+	+			+	+	+	+
	AF13+O	SDmPGVLR ⁺ Fa	+		+	+	+	+	+	+
	AF10+O ₂	GFGDEmSmPGVLR ⁺ Fa			+		+	+	+	+
	AF10+O	GFGDEMsmPGVLR ⁺ Fa	+	+			+	+	+	+
	AF10+O ₂	FGDEmSmPGVLR ⁺ Fa	+	+						
	AF10+O ₂	GDEmSmPGVLR ⁺ Fa	+			+	+	+	+	+
	AF10+O ₂	DEmSmPGVLR ⁺ Fa					+	+	+	+
	As-NLP-13-2+O	SNAFDRNFmNFa	+	+	+	+			+	+
	As-NLP-13-6+O	SDAFSRNFmNFa	+	+	+	+				
<i>afp-5</i>	As-NLP-13-3	ESQFSRDLNFa	+	+	+	+				
	AF17+O	DFDRDFmHFa	+			+				
	AF22	NGAPQPFVR ⁺ Fa	+	+						
	PepSE	SGRVDHHDILSTLQRLQLANE			+	+			+	+
<i>afp-11</i>	AF29	NAEPNLR ⁺ Fa	+	+	+	+	+	+		
	AF11	SDIGISEPNFLR ⁺ Fa	+	+			+	+	+	+
	AF28	SAEPNLR ⁺ Fa	+		+	+	+	+	+	+
	AF32	GSDPNFLR ⁺ Fa	+		+		+	+	+	+
	AF27	PADPNFLR ⁺ Fa					+	+	+	+
	AF36+O ₂	VPSAADmmIR ⁺ Fa	+	+	+	+	+	+	+	+
	AF19	AEGLSSPLIR ⁺ Fa	+	+	+	+	+	+	+	+
<i>afp-13</i>	AF34+O	DSKLmDPLIR ⁺ Fa	+	+	+	+				
	AF34+O	LmDPLIR ⁺ Fa					+	+	+	+
	PepTT	TPPEEDLLGRFT	+	+	+	+				
	Pep(TL+O	TNImGENRLNRNL							+	+
<i>As-nlp-3</i>	As-NLP-3-1	snfdgreasprf	+	+	+	+				+

Gene	Peptide	Sequence	Run1			Run2			Run3			Run4		
			M	S		M	S		M	S		M	S	
	As-NLP-3-2	AINPFTDSIa												
	As-NLP-3-4*	NTGIAAERDLPKRYFDALAGQSLa				+						+		
	As-NLP-3-4	YFDALAGQSLa									+			
As-nlp-5	As-NLP-5-2*	TSISRSKDAKKPSLDILEGAGFSPL												+
	As-NLP-5-4+O	ALDDLEGVGFGGmL												
As-nlp-7	As-NLP-7	NHLIGFDDPRLFSSSYa				+								
As-nlp-8	As-NLP-8-2*	SFDRIDGSAFGPHRH				+						+		+
	As-NLP-8-5	AFDRIEGSGFGLD										+		
	As-NLP-8-4	AFDRIEGAGFGLS										+		
	As-NLP-8-4-	AFDRIEGAGFG										+		
As-nlp-9	As-NLP-9-2	FGSFSPIYFYQQPK				+						+		
As-nlp-15	As-NLP-15a-3	SLDSPANRGFTTFD				+						+		
	As-NLP-15b-2-	SFDSLNSRGFTG										+		
	As-NLP-15b-3-	AFDSLVGHGFTG										+		
As-nlp-21	As-NLP-21-2	GGGRSFRLLSLE												+
As-nlp-23	As-NLP-23-2	SLALGRYDFRPG				+								+

* extended form of peptide; - truncated form of peptide; m oxidized methionine; a amidated terminus; novel peptides in bold.