1   # A combined strategy of neuropeptide predictions and tandem

2   mass spectrometry identifies evolutionarily conserved ancient

3   neuropeptides in the sea anemone *Nematostella vectensis*

4

5   Eisuke Hayakawa[1, 2]*, Hiroshi Watanabe[2, 3], Gerben Menschaert[4], Thomas W. Holstein[3],

6   Geert Baggerman[4,5] and Liliane Schoofs[1]

7   [1] Research Group of Functional Genomics and Proteomics, KU Leuven, Leuven, Belgium

8   [2] Evolutionary Neurobiology Unit, Okinawa Institute of Science & Technology, Okinawa, Japan

9   [3] Centre for Organismal Studies (COS), Heidelberg University, Heidelberg, Germany

10   [4] Faculty of Bioscience Engineering, Laboratory for Bioinformatics and Computational Genomics, Ghent

11   University, Ghent, Belgium

12   [5] CFP/Ceproma, University Antwerpen, Antwerpen, Belgium

13   [6] VITO, Applied Bio & molecular Systems (ABS), Mol, Belgium

14

15   *Corresponding author

16   E-mail: eisuke.hayakawa@oist.jp

17

Abstract

19 Neuropeptides are a class of bioactive peptides and are responsible for various physiological

20 processes including metabolism, development and reproduction. Although accumulated genome and

21 transcriptome data have reported a number of neuropeptide candidates, it still remains difficult to obtain a

22 comprehensive view of neuropeptide repertoires due to their small and variable nature. Neuropeptide

23 prediction tools usually work only for those peptides for which sequentially related homologs have

24 previously been identified. Recent peptidomics technology has enabled systematic structural identification

25 of neuropeptides by using the combination of liquid chromatography and tandem mass spectrometry.

26 However, obtaining reliable identifications of endogenous peptides is still difficult using a conventional

27 tandem mass spectrometry-based peptide identification approach using protein database because a large

28 search space has to be scanned due to the absence of a cleavage enzyme specification. We developed a

29 pipeline consisting of the prediction of *in silico* cleaved endogenous neuropeptides followed by peptide-

30 spectrum matching enabling highly sensitive and reliable neuropeptide identification. This approach

31 effectively reduces the search space of peptide-spectrum matching, and thus increases search sensitivity. To

32 identify neuropeptides in *Nematostella vectensis,* a basal eumetazoan having one of the most primitive

33 nervous systems, we scanned the *Nematostella* protein database for sequences displaying structural

34 hallmarks of metazoan neuropeptides, including C/N-terminal structures and modifications. Peptide-

35 spectrum matching was performed against the *in silico* cleaved peptides and successfully identified dozens

36 of neuropeptides at high confidence. The identification of Nematostella neuropeptides structurally related

37 the tachykinin, GnRH/AKH, neuromedin-U/pyrokinin peptide families indicate that these peptides already

38 originated in the eumetazoan ancestor of all animal species, most likely concomitantly with the

39 development of a nervous system.

40

41

Introduction

43 Neuropeptides are a highly diverged group of messenger molecules and function as essential

44 communicators for fairly standard physiological processes such as muscle contraction, food digestion,

45 growth, development, reproduction as well as for more complex actions and long-term effects such as

46 behavioural adaptation, memory formation and even ageing. A neuropeptide is usually encoded in a larger

47 neuropeptide precursor gene that also typically encodes an N-terminal signal peptide. The precursor is

48 translated on the rough endoplasmic reticulum, and the signal peptide is removed by a signal peptidase.

49 Afterwards, processing enzymes in the immature secretory granules typically produce a mature

50 neuropeptide by processing the precursor at cleavage sites. A number of cleavage enzymes with specific

51 recognition patterns have been currently identified [1,2]. After cleavage, the cleaved peptides can undergo

52 various post translational modifications (PTMs) [3]. The N-terminal or C-terminal residues are often

53 modified and these modifications often contribute to the stability of the mature peptides. The most

54 frequently observed modification is amidation at the C-terminus. It occurs in many neuropeptides and is

55 often required for their biological activities. Enzymatic transformation of the C-terminus from a glycine to

56 an alpha-amide is catalyzed by peptidylglycine alpha-amidating monooxygenase [4]. First, peptidylglycine

57 is transformed into peptidyl-alpha-hydroxyglycine in the presence of copper, ascorbate, and molecular

58 oxygen and then converted to peptide alpha-amide and glyoxylate. C-terminal amidation protects the

59 mature peptide from enzymatic degradation from C-terminus. Pyroglutamic acid is widely observed at N-

60 termini of various neuropeptides and protects the peptide chain from enzymatic degradation from N-

61 terminus. In addition to PTMs, many bioactive peptides contain a proline residue in the second or third

62 position from the N-terminus. This feature also protects the peptide from peptidase activity at the N-

63 terminus [5]. All these structural characteristics can be widely observed in neuropeptide sequences across

64 the Animal Kingdom [6].

65 Recent advances in mass spectrometry and liquid chromatography technology have led to the

66 establishment of peptidomics, an efficient tandem mass spectrometry-based approach to identify

67 neuropeptides. The combination of tandem mass spectrometry and peptide-spectrum matching tools

68 enables systematic neuropeptide identification [7–9]. Unlike neuropeptide prediction tools that are based on

69 sequence similarities, this peptidomics approach allows providing evidence for the *in vivo* presence of the

70 cleaved peptides, in addition to determining the presence of PTMs of the peptide sequences [10,11].

71 However, in search of endogenous small peptides using conventional peptide-spectrum matching tools,

72 achieving reliable identifications still remains difficult. Unlike proteomics, in which *in vitro* generated

73 enzymatic digests are identified, cleavage sites in a precursor of endogenous neuropeptide cannot readily be

3

74   accurately predicted. Therefore peptide-spectrum matching has to be performed without any enzyme

75   specification, leading to a huge search space often resulting in poor identification confidence values.

76   Besides, peptide-spectrum matching has to be done taking PTMs into consideration, even further increasing

77   the search space [12].

78   Another and rather commonly employed approach to search for neuropeptides are homology-based

79   database searches. Recent strides in sequencing genome and transcriptome in a wide variety of animal

80   species have predicted bioactive peptides *in silico* [13–16]. Several researches successfully track down the

81   evolutionarily conserved neuropeptides within a phylum or between closely related phyla, by means of

82   sequence homology searching against protein sequence datasets [17]. This approach is very useful for

83   searching peptide sequences that have been evolutionarily conserved between closely related species. .

84   However, when peptides have become evolutionarily diverged, such as is the case for evolutionarily

85   ancient organisms, the homology-based prediction of new neuropeptides may not always be successful. On

86   top of this difficulty, within a particular peptide sequence, only a short motif required for the peptide's

87   biological activity, has in general been conserved during evolution [16,18] whereas the sequence of the

88   none-peptide coding region of the precursor, is much less or not at all conserved. All these issues hamper

89   reliable sequence similarity-based peptide prediction. Especially for evolutionarily distant species, it

90   remains difficult to identify neuropeptides and corresponding genes based on homology searches and

91   experimental validation is required in these cases. To deal with these hurdles, we have developed an

92   alternative strategy to identify neuropeptides with high sensitivity and confidence. The strategy is a

93   combination of the reduction of protein datasets to potential mature neuropeptide structures predicted based

94   on the typical hall marks(signal peptide, specific cleavage sites, PTMs), combined with peptide-spectrum

95   matching. The first step is effective in narrowing the search space for the peptide-spectrum matching and as

96   such enables high sensitive peptide identifications. Based on this strategy, we here present a systematic

97   identification of neuropeptides from *Nematostella vectensis*, one of the most evolutionarily ancient animals

98   with a nervous system. We successfully identified 20 neuropeptides, many of which have not been

99   predicted or annotated as neuropeptides when the genome of this organism was sequenced [19]. Compared

100   to a conventional peptidomics workflow, this approach enables the identification of neuropeptides with the

101   highest sensitivity.

102

103

Materials and Methods

Sample preparation

106   Adult polyps of *Nematostella vectensis* were kept in 1/3 artificial seawater at 18 °C and fed three

107   time per week with nauplius larvae of *Artemia salina*. The culture medium was changed once a week. Five

108   adult polyps were homogenized with a probe sonicator in 1.5 ml methanol/water/formic acid (FA) solution

109   (90:9:1) in 3 cycles (on for 5 seconds and off for 5 seconds) on ice. Large proteins were removed by

110   centrifugation at 10,000 xg for 15 minutes, and supernatant was transferred to new tube. The sample was

111   freeze-dried using a vacuum centrifuge (Speedvac concentrator SVC200H, Savant, USA) and stored at -

112   80°C for further treatment.

MALDI-TOF/TOF mass spectrometry analysis

114   The sample was prefractionated on a C18 column (BEH C18 column, Waters, Milford, MA, USA)

115   using mobile phase at high pH (MilliQ and acetonitrile with ammonium hydroxide (20 mM), pH 10) and

116   five fractions were made during the gradient (B: 5 to 90 % in 30 minutes) at a flow rate of 100 μL /min.

117   Prefractionated samples were further separated on C18 column at low pH. The eluent was (A) water

118   containing 0.5% FA and (B) 90% acetonitrile in 0.5% aqueous formic acid. The column was first washed

119   and equilibrated with eluent A and 5% of eluent B. After loading the sample, a linear gradient from 5% B

120   to 60% B in 60min at a flow rate of 100 μL /min was used as the mobile phase. Thirty fractions of 200 μL

121   were collected from the beginning of the gradient using an automatic fraction collector. The resulting

122   samples were dried in a Speed-vac-and stored at -80°C until further analysis. Fractionated samples were

123   resuspended in 1.5 μL water/ acetonitrile/FA (50/49.5/0.5 v/v). Subsequently they were transferred onto a

124   MALDI target plate (Bruker Daltonics, Bremen, Germany) and mixed with 1.5 mL of a saturated solution of

125   CHCA in 50% acetonitrile containing 0.5 % FA. After evaporation of the solvent, the MALDI target was

126   introduced into the mass spectrometer ion source. Tandem mass spectrometry analysis   were performed on

127   an Ultraflex II instrument (Bruker Daltonics, Bremen, Germany) in positive ion, reflectron mode. The

128   instrument was calibrated externally with a commercial peptide mixture (peptide calibration standard,

129   Bruker Daltonics). All spectra were obtained using Flex Control software (Bruker Daltonics, Bremen,

130   Germany). The plate was initially examined in MS1 mode and spectra were recorded within a mass range

131   from m/z 500 to 4000. Subsequently, the peaks with S/N 10 were selected and used for the optimized LIFT

132   method from the same target. All tandem mass spectra were processed by means of the FlexAnalysis

133   software (Bruker Daltonics, Bremen, Germany), and m/z values and intensities of each peak were recorded in

134   peak list files.

135

Mass spectrometry data analysis

137   In this study, potential neuropeptide sequences are systematically extracted from proteins in order

138   to reduce the search space in peptide-spectrum matching as follows. Amino acid sequences of proteins

139   predicted from the genome of *Nematostella vectensis* were obtained from JGI (*Nematostella vectensis*

140   version 1.0, all models) [20]. An in-house software module was developed enabling the prediction and

141   extraction of potential mature neuropeptide sequences from the *Nematostella* proteins. First, the amino acid

142   sequence of a protein is scanned for residues corresponding to input amino acid pattern (e.g. GK/R for C-

143   terminal amidation) in a sliding window fashion. Secondly, an amino acid motif corresponding to an N-

144   terminal structure (e.g. X(X)P) is searched at the N-terminal side from C-terminal motif found within a

145   given amino acid length. In case multiple locations found to match N-terminal motif, all of them are

146   considered to be potential N-termini. All combinations of input amino acid patterns are applied. All amino

147   acid sequences that meet the criteria are cleaved *in silico* and written in a FASTA formatted file and used as

148   target database for peptide-spectrum matching. For the present study, the following amino acid motifs were

149   used to extract potential neuropeptide precursor protein sequences.

150

151   C-terminal motif

152   G[K/R]↓ : glycine loss-amidation

153   N-terminal motif

154   ↓XP: proline from the second position from N-terminus

155   ↓XXP: proline from the third position from N-terminus

156   ↓Q: Glutamate for the formation of pyroglutamic acid

157   (↓ indicates the cleavage site)

158   The software was coded in C++ using Visual Studio 2008 (Microsoft) and Boost library. The software is

159   available upon request.

160

161   All tandem mass spectra were searched against a database holding (A) the entire proteins and (B) the

162   predicted neuropeptide sequences, using the Mascot search engine version 2.3 (Matrix science, London,

163   UK). Peptide-spectrum matching were performed allowing variable modifications of glycine-loss amidation,

164   pyroglutamic acid (Glu), and oxidation (Met), which are common post-translational modifications of

165   neuropeptides. Searches against the predicted neuropeptide sequence dataset were performed with "No

166   Cleavage" setting, in which cleavage of input sequences is not considered and precursor masses are

167   matched to the theoretical masses of intact peptide sequences. Mass tolerance for precursors and fragment

168   ions were set to 0.4 and 0.8 Da. In this study, peptides were first tentatively identified with a Mascot expect

169   value less than 0.05 and then further confirmed by manual verification of the product ions assigned. Decoy

170   search was conducted by Mascot decoy database function. Signal peptides of the precursors of identified

171   peptides was examined using SignalP version 4.1 [21] .

172

Expression analysis

174   Partial ORF sequences for the identified peptides were obtained from the National Center for

175   Biotechnology Information (NCBI) trace archive of the *Nematostella vectensis* (data generated by the Joint

176   Genome Institute) and from *Stellabase* (http://cnidarians.bu.edu/stellabase/index.cgi). Gene-specific

177   primers were designed based on the ORF sequences. The primer sequences are available upon request. PCR

178   products for the peptide genes were subcloned into pGEM-T (Promega) and sequenced. Riboprobes were

179   synthesized and purified as described previously [22]. *In situ* hybridization was performed as described

180   previously [23] with following modifications: specimens were fixed with 4% paraformaldehyde/PBS +

181   0.1% Tween 20 (PBST) for 1 hr, washed with methanol for 3 times and stored at -20ºC. Hybridization of

182   0.4- to 1.1-kb digoxygenin (DIG)-labeled antisense RNA probes was carried out using hybridization

183   solution containing 1% SDS at 50-65ºC for at least 22 hrs. For post-hybridization washes, specimens were

184   washed by serial dilutions (75%, 50%, 25%) of hybridization solution with 2× SSC at 55ºC. After DIG-

185   labeled probe was visualized using BM purple (Roche), specimens were washed with PBST.

7

186

187

Results

Strategy

190   Systematic neuropeptide identification was performed for *Nematostella vectensis*, using a

191   combination of neuropeptide prediction combined with tandem mass spectrometry. Fig 1 shows an

192   overview of the used approach. Protein database of *Nematostella vectensis* were first processed with our

193   newly developed neuropeptide prediction tool to extract potential neuropeptide sequences based on their

194   structural hallmarks such as cleavage sites and N/C-terminal PTMs . Fragmentation spectra were acquired

195   by means of LC-MALDI TOFTOF. Unlike conventional peptide-spectrum matching   for endogenous

196   peptides or enzymatic digests, we narrowed the search space and only used the sequences of *in silico*

197   cleaved potential neuropeptides a target sequence dataset for peptide-spectrum matching. Typically in

198   peptidomics, searches are often done considering any amino acid residue as potential cleavage sites. In the

199   present method, the search is done directly against the mature forms of potential neuropeptide sequences in

200   a manner similar to top-down analysis of protein .

201

202

203

204   **Fig 1. Scheme of the neuropeptide identification strategy using a combination of *in silico* cleaved**

205   **neuropeptide predictions and peptide-spectrum matching.** The N/C-terminal motifs are used to scan the

206   protein database for the candidates of neuropeptides. The detected sequences are exported as a target

207   database for top-down MS/MS search.

208

Building prediction models

209

210   To find common structural hallmarks of neuropeptides in cnidarian species, we collected

211 neuropeptide sequences from the UniProt sequence database and other publically available resources

212 (supporting table 1)[24–33]known neuropeptides in cnidarian species, 52 % have N-terminal prolines and

213 54 % have N-terminal glutamines, which are used for the conversion to pyroglutamic acid. In total, 84 % of

214 cnidarian neuropeptides have either proline or glutamine at N-termini, indicating these structures can be

215 hallmarks to effectively extract potential neuropeptide sequences from protein sequences. All of the known

216 cnidarian neuropeptides contained an amidation at their C-termini, which is a commonly observed PTM for

217 neuropeptides in all species across the Animal Kingdom. As indicated by many studies, cleavages at

218 dibasic residues ([KR] X{0,2,4,6}[KR]), which is the target site for the prohormone convertases, is

219 commonly used to generate neuropeptides from precursor proteins.[1] This processing pattern can be found

220 in peptide precursor proteins of many bilaterian species.[9] It should be noted, however, that in cnidarians

221 dibasic residues are rarely used as cleavage sites   within neuropeptide precursors (supplemental table 1).

222 They do not show particularly conserved cleavage sites. Based on these observations, we selected N-

223 terminal protective structures (Q or E for pyroglutamic acid formation or P at the second or third position

224 from the N-termini) as well as C-terminal amidation (glycine and basic residues (K or R) as structural

225 hallmarks for neuropeptide prediction. This is because these structures can be found in the vast majority of

226 neuropeptides in cnidarian species and are not limited to particular neuropeptide families (supplemental

227 table 1), thereby enabling comprehensive and untargeted neuropeptide identification. This information

228 (types of amino acid residues and locations on peptides) was fed to our neuropeptide prediction tool as

229 input, and protein database of *Nematostella* were scanned for potential neuropeptide sequences using the

230 combination. Without any cleavage site specification (conventional peptidomics search), about 100,000 to

231 200,000 peptide sequences were found to fit the molecular masses of each observed peptide in the sample

232 with the given mass tolerance in the *Nematostella* proteins. By comparison, in the newly constructed

233 database that contained only *in silico* cleaved predicted neuropeptide sequences, only 400 to 800 peptides

234 could be found. This efficiently reduced the size of the search space.

235   The tandem mass spectra were acquired from off-line LC-MALDI TOF-TOF analysis of a peptide

236 extract of *Nematostella vectensis*. Peptide identification was carried out by using the Mascot peptide-

9

237   spectrum matching tool, taking into account pyroglutamic acid formation, glycine-loss amidation and

238   oxidation on M. The dataset of extracted potential neuropeptide sequences was used as target sequence

239   dataset for peptide-spectrum matching, and the search was done in the top down fashion with "no-

240   cleavage" setting. Twenty unique peptides were identified with a Mascot e-value < 0.05. Table 1 shows the

241   identified peptide sequences. The amino acid sequences and the details of precursor protein coding genes

242   can be found in supplemental table 2.

243

Family name	Sequence	Mascot e-value		
		predicted peptides	protein database	basic cleavage sites
HIRamide	APPLDLSGPAYFHIRa	**1.9E-02**	4.2E+00	nd
	GPPYIDLTEPSFFHIRa	**8.8E-09**	**2.1E-06**	nd
	NPPIDLGPAYFHIRa	**7.3E-07**	**1.6E-04**	**5.9E-07**
	pQPPIDLSPAAYFHIRa	**1.1E-04**	**2.5E-02**	**7.7E-05**
	pQPPLDLGPAYFHIRa	**5.0E-07**	**1.1E-04**	nd
	pQPPYLDLGEPSFFHIRa	**4.5E-04**	1.0E-01	**3.6E-04**
	pQPPYLDLTPAYFHIRa	**2.5E-04**	5.5E-02	nd
	pQPPYLDLTPSYFHIRa	**5.4E-06**	**1.2E-03**	nd
	pQPPMIDLSEPAFFHIRa	**1.2E-06**	**2.7E-04**	nd
	pQQPPMIDLSEPAFFHIRa	**1.7E-03**	3.5E-01	nd
PRGamide	GPRGGRATEFGPRGa	**1.2E-04**	**2.8E-02**	nd
	GPRGGREVNLEGPRGa	**6.7E-03**	1.7E+00	nd
QWamide	IPPQGFRFNQWa	**3.4E-02**	8.2E+00	nd
	IPPQGLRFNQWa	**2.9E-03**	6.7E-01	nd
	IPPQGLRFSQWa	**2.8E-03**	6.4E-01	nd
RFamide	MPEQDANPQTRFDa	**2.3E-05**	**5.3E-03**	nd
	pQGRFGREDQGRFa	**3.2E-03**	6.3E-01	nd
RPamide	FPPGFHRPa	**2.0E-03**	4.9E-01	nd
RHamide	GPPMIKIPVRHa	**2.0E-03**	5.0E-01	nd
	GPPMoIKIPVRHa	**7.7E-04**	2.5E+01	nd

244   **Table 1.** Sequences of detected peptides and their Mascot E-value. E-values lower than the threshold (0.05)

245   were indicated in bold. C-terminal amidations, oxidation, and N-terminal pyroglutamic acids are

246   respectively indicated as "a", "o" and "p".

247

248   We compared the peptide identification performance of our approach with the conventional

249   peptide-spectrum matching against the whole protein dataset in which every amino acid residue is

250   considered as a potential cleavage site. As shown in table 1, the e-values of the search results using our

251   predicted peptide dataset were significantly improved and our approach yielded more than twice as many

252   peptide identifications. Fig 2 shows the overall distribution of the e-values of these peptides in the target

253   and decoy searches. The e-values of the peptide hits in the protein database (Fig 2C) are distributed in the

254   higher range and many of them show similar e-values as the decoy search result (Fig 2D), making the

255   peptide sequence identification difficult. On the other hand, the e-values of the peptide hits using the

256   reduced database of predicted candidate peptide sequences (Fig 2A) are distributed in a lower range, and

257   show clear separation from the range of e-values of the decoy search result (Fig 2B).

258

259   **Fig 2. The distribution of e-values of the hits found in the datasets.** Each circle indicates the e-value of
260   the top-scoring hit in the dataset. A: Search against the predicted neuropeptide dataset. B: decoy of the
261   predicted neuropeptides dataset. C: the whole protein models. D: decoy of the whole protein models.

262

263   In addition, to demonstrate the difference from the method in which only basic cleavage sites are

264   used to narrow search space [34], we also scanned the protein database using the cleavage rule of

265   prohormone convertases ([KR] X{0,2,4,6}[KR]). Among the peptides detected by using the structural

266   hallmarks of neuropeptides, only 3 were identified (Table 1). This indicates that, narrowing down search

267   space with a simple cleavage rule failed to predict neuropeptide sequences.. The peptide sequences and

268   their precursors acquired in this study indicate that typical prohormone convertase cleavage sites (K/R) are

269   not commonly utilized for the neuropeptide precursor processing in *Nematostella*, which agree with the

270   neuropeptides found in other cnidarian species[25,30,35–37].

271 # Novel neuropeptides in *Nematostella vectensis*

272   Most of the peptides identified in the present study have not been reported yet, and their precursor

273   protein-coding genes were not annotated as neuropeptide precursor, except for the RFamide precursor. Fig

274   3 shows the overview of the precursors that contain identified peptides. Overall, these precursors contain

275   multiple copies of the mature peptide sequences sharing the same C-terminal structures. It should be noted

276   that the precursors also contained multiple other potential neuropeptide sequences which contains same

277   structural hallmarks used to the *in silico* cleavage, but not observed in this study. It is unclear whether those

278   peptides are not produced or because of the sensitivity of mass spectrometry analysis is insufficient to

279   detect those peptides.

280

281

282   **Fig 3. Overview the structures of neuropeptides precursor proteins detected in this study.** Gray

283   indicates the signal peptide. Black indicates the locations of the detected and potential mature

284   neuropeptides.

285

286   The HIR peptide precursor contains 9 structurally related peptide sequences and a typical signal

287   peptide sequence at the N-terminus. They share a high degree of structural similarity at the C-terminus with

288   Arthropoda Tachykinin peptides (Fig 4A). Especially, *Locusta*-tachykinin-1[38] displays strong similarity

289   to the *Nematostella* HIR peptides. Tachykinin and related peptides are well studied neuropeptides

290   conserved in both invertebrates and vertebrates. Based on the C-terminal motifs, Tachykinin and related

291   peptides can be further classified into two subgroups: the "-FXGLMa" and "-GFXGXRa" subfamilies.

292   *Nematostella* HIR peptides show similarity to the "GFXGXRa" subfamily; especially the presence of

293   phenylalanine or tyrosine residues, and C-terminal arginine is conserved in *Nematostella*. Arthropod and

294   Nematostella share the PXXFYXXRamide motif.

295

296

297   **Fig 4. Structural similarities of identified neuropeptides in *Nematostella vectensis* and other species.**

298   A: HIRamides, Tachykinin related peptides[16,39–43] . B: PRGamides and PRXamide related

299   peptides[44–48] . C: QWamides, AKH and Hypertrehalosaemic hormones [49–54]. Conserved amino acid

300   residues are shown in red.

12

301

302   RFamides, a neuropeptide family widespread among bilaterian organisms and also previously

303   reported in cnidarians including sea anemone, were also detected in this study. In this study an additional

304   member of RFamide MPEQDANPQTRFDa was identified in the precursor protein of RFamide. This new

305   member has an additional D at C-terminus, although all the other members are showing the RFamide motif.

306   Two peptides, GPRGGRATEFGPRGamide and GPRGGREVNLEGPRG, share the C-terminal motif

307   GPRGamide. Their peptide precursor contains a large number of peptide copies all sharing this motif. The

308   peptides share sequence similarities with protostomian pyrokinins that are characterized by the C-terminal

309   sequence FXPRLamide [55,56]. Pyrokinins are involved in the stimulation of gut motility, the production

310   and release of sex pheromones, diapause, pupariation and feeding. (Fig 4B) [57] [58].

311   Three peptides derived from a single precursor protein display the C-terminal sequence motif

312   FSQWamide or FNQWamide (Fig 4C). This motif aligns with the motif that typifies the bilaterian

313   AKH/GnRH peptide family which is known to regulate energy metabolism and reproduction [59] ..

314   In addition to the peptides described above, we discovered two groups of peptides (RPamide and

315   RHamide) originated from two precursor proteins. We could not find any homologies to known

316   neuropeptide families (table 1).

317

318   # Expression patterns of HIRamide, PRGamide, VRHamide

319   # neuropeptides

320   Our whole mounts in situ hybridization (WISH) analyses have confirmed that the peptide genes

321   encoding the identified peptides (HIRamide, PRGamide, VRHamide) are exclusively expressed in neurons.

322   As shown in fig 5, the expression of all peptide genes that could be detected by WISH was observed in

323   specific cells housed mainly in the endodermal layer at the juvenile polyp stages. Careful observations at

324   the cell morphology demonstrated that the cells exhibiting the expression of the genes shows a round-

325   shaped cell body with neurite-like processes (Fig 5, lower panel). This indicates that these detected peptides

326   are in fact expressed in neurons, and are neuropeptides. WISH experiments demonstrated also that the

327   HIRamide and PRGamide neuropeptides are strongly expressed around the mouth opening (Fig 5a). This

328   finding indicates that neuronal subsets expressing these neuropeptides develop at the oral side to form

1

329 region-specific neural network. The tissue around mouth of cnidarian polyps have been shown to express a

330 bunch of neuronal markers including RFamide, which is also evolutionarily conserved among metazoans

331 (Fig 5, upper panel) [19], and to develop a regionalized nervous system which is known as the oral nervous

332 system (*Nematostella*) or the nerve ring (*Hydrozoa*) [60–62]. Expression pattern of VRHamide encoding

333 gene showed a sharp contrast to that of HIRamide, PRGamide and RFamide genes. The expression level of

334 VRHamide gene at the oral tissue was under detection limit. Instead, it was strongly and exclusively

335 detected in neurons located at the most distal region of the tentacle endoderm (Fig 5, upper panel). This

336 unexpected and interesting expression pattern suggests the specific function of the VRHamide for

337 development and/or contractile response of the tentacles (see Discussion).

338

339   **Fig 5. WISH staining of juvenile polyps of *Nematostella vectensis* (8 days post fertilization) showing**

340   **localized expressions of *HIRamide, PRGamide RFamide* and *VRHamide* genes at low magnification**

341   **(upper panels) and high magnification (lower panels).** Scale bars, 100 μm (upper) and 50 μm (lower).

342

Discussion

344   We studied the peptidome in *Nematostella vectensis*, which belongs to the Cnidaria that have one

345   of the most basal nervous systems in the Animal Kingdom. Recent advances in mass spectrometry

346   technologies have greatly enhanced the identification of neuropeptidomes from biological matrices [12,63–

347   65]. However, an effective informatics solution for neuropeptide identification is still missing and leaving

348   the interpretation of tandem mass spectral data still difficult. We overcame this difficulty by performing

349   peptide spectrum matching against a database comprised of *in silico* cleaved putative predicted

350   neuropeptide sequences. We successfully identified 20 peptides encoded in six neuropeptide precursor

351   genes from *Nematostella* and confirmed experimentally that at least four precursor genes are expressed in

352   neurons exhibiting distinct neurophysiological activities. Our approach can be easily incorporated with

353   recent advanced technologies in peptidomics, and such a combination will greatly improve the analysis of

354   the neuropeptidome in a broad range of biological materials.

Strategies for the improvement of peptide identification

356   Peptide-spectrum matching has become a main tool for endogenous peptide identification. Unlike

357   classical methods such as Edman degradation, this approach takes advantage of the high sensitivity of

358   modern mass spectrometry instruments and high throughput peptide search engines and therefore has

359   enabled comprehensive neuropeptide characterization. One of the main difficulties in peptidomics is that

360   the peptide-spectrum matching needs to be done with the "no enzyme" setting, which results in an

361   enormous search space. Consequently the peptide-spectrum matching result is regularly worse as compared

362   to a peptide search in a proteomics setting.

363   The peptide characterization strategy employed in this study is a combination of predicting

364   potential neuropeptide sequences and peptide-spectrum matching. This approach is effective in narrowing

365   the search space and thereby significantly increases the sensitivity of peptide identification. Fälth *et al.*

1

366   employed a similar approach to narrow search space in order to improve peptide-spectrum matching [34].

367   They used the most common cleavage site rule to extract potential neuropeptide sequences from a large

368   protein database, and significantly improved peptide identification rate. Southey *et al.* developed a

369   dedicated tool to predict the basic cleavage sites in neuropeptide precursor genes.[66] The prediction

370   engine, NEUROPRED, needs to be trained on the information of known neuropeptides and their

371   precursors; hence a large number of identified peptides and the information of processing sites are required.

372   Compared to the prediction with simple motif matching, the prediction by NEUROPRED using the logistic

373   regression modelling is more reliable, but on the other hand rather strict and potentially discards many

374   forms of cleaved peptides especially if the large amount of training data is not available. These approaches

375   are effective for the prediction of neuropeptides in well-studied animal phyla. On the other hand, it's

376   difficult to predict peptides from less-well studied animal models such as cnidarians. Unlike bilaterian

377   animals, non-basic cleavage sites are commonly observed in cnidarian neuropeptide precursors

378   [25,27,31,33,35–37,67]. In fact, most of the peptides identified in the present study are generated via

379   cleavage at non-basic residues and unconventional sites. Therefore, the approach we took in the present

380   study, i.e. the prediction of putative neuropeptide sequences based on structural hallmarks such as

381   protective N-terminal pGlu or X(X)P and C-terminal amidation is more comprehensive and effective to

382   preserve as many potential neuropeptide sequences as possible as the target database for peptide-spectrum

383   matching. Evidently, the approach has also a drawback as it only allows identification of all neuropeptides

384   that display these specific features. Non-amidated peptides or peptides that are not N-terminally protected

385   are not in the database. At this moment, there are no means other than homology-based searches, or

386   Neuropred prediction tools to search for the latter peptides, and thus far these tools are ineffective for

387   cnidarian neuropeptides because of aforementioned reasons.

Neuropeptide evolution

389   Many pioneer studies in the late 80ties and early 90ties already pointed to homologies between

390   protostomian and deuterostomian neuropeptides based on neuropeptide sequence identifications in

391   vertebrate, insect, molluscan and cnidarian species [18,30,68–72]. However, until today knowledge on

392   signalling molecules, including neuropeptides, in eumetazoan animals is scarce. A previous study showed a

393   highly conserved set of genes in the *Nematostella vectensis* genome, including molecules involved in

394   neurotransmission [19,73]. The genome displays a number of receptor genes such as G-protein coupled

395   receptors showing high degrees of similarities to known neuropeptide receptors. It is therefore highly likely

396   that this primitive animal contains a variety of signalling peptides. Anctil *et al.* has reported potential

397   neuropeptide coding genes by using homology searching [74]. Although those predicted precursor proteins

398   contain short repeats of neuropeptide-like motifs, it is difficult to prove their identities as these

399   neuropeptide precursors were predicted solely based on their protein sequence without any experimental

400   evidence. In fact some of the predicted peptide precursor proteins in this study show a significant degree of

401   structural similarity to non-neuropeptide precursor proteins, making it difficult to decide whether they are

402   true neuropeptide precursors. In the present study we were unable to identify any of those predicted

403   peptides, either by conventional "no-enzyme" searches nor by searches against the predicted neuropeptides.

404   Homology-based searches based on short neuropeptide sequences is thus still difficult and, more

405   importantly, the homology-based prediction of neuropeptide precursor genes does not provide evidence for

406   the occurrence of predicted neuropeptides *in vivo*. Therefore, peptide identification approaches that are

407   based on conventional biochemical purification methods and powered by upcoming bioinformatic

408   technologies are necessary to empirically prove the predicted peptides and will be important to identify not

409   only evolutionary conserved peptides but also evolutionarily derived or species-specific peptides that

410   correspond to the diverged physiological traits.

411   Our data have also important implications on our insight into the evolution of neuropeptides. Since the

412   neuropeptide repertoire in ancient eumetazoans is largely unknown, the origin of most of neuropeptides

413   found in bilaterian animals remains difficult to reveal. Jekely (2013) performed a similarity-based

414   clustering analysis of genes encoding neuropeptides and neuropeptide GPCRs across metazoan phyla and

415   concluded that the last common ancestor of eumetazoans had various small amidated peptides including

416   RFamide, RYamide, and Wamide) [75]. Ancestral urbilaterian neuropeptide-receptor families include

417   GnRH, vasopressin, GnIH/SIFamide, CRF/diuretic hormone, calcitonin/DH31, NPY/NPF, neuromedin-

418   U/pyrokinin, CCK/sulfakinin, galanin/allatostatin-A, and orexin/allatotropin [75,76]. It was suggested that

419   these neuropeptide families may have originated concomitantly with the origin of a complex bilaterian

420   body plan having a through gut with novel controls for food intake and digestion, excretory and circulatory

421   systems, light-controlled reproduction, a centralized nervous system, complex reproductive behavior, and

422   learning.

1

423   The present study indicates that the extent of the conservation of some neuropeptide families is

424   even deeper as proposed in aforementioned studies, as we were able to identify neuropeptides of the

425   GnrH/AKH, tachykinin and neuromedin-U/pyrokinin families in a cnidarian Nematostella, suggesting that

426   these neuropeptide families were already present in the common ancestors of all eumetazoan species.

427   Because *Nematostella vectensis* is a member of the Anthozoa, the most basal cnidarian group, the

428   neuropeptides identified in the present study might constitute some of the most ancient neuropeptide in

429   animal evolution. At least the tachykinin, GnRh, neuromedin-U and RFamide neuropeptides may have

430   originated concomitantly with the formation of a nervous system. All these neuropeptides play a variety of

431   roles in animal physiology, including feeding and metabolism, locomotion, reproduction, homeostasis,

432   biorhythms and associated behaviors. We anticipate that the discovery of the receptors, downstream targets,

433   and functions of cnidarian representatives of ancient eumetazoan neuropeptide families will provide

434   important new insights into the evolution of neuropeptide function in the Animal Kingdom

435

Acknowledgement

439

References

441   1.   Fricker LD. Neuropeptide-processing enzymes: applications for drug discovery. AAPS J. 2005;7: E449-

442   55. doi:10.1208/aapsj070244

443   2.   Hook V, Funkelstein L, Lu D, Bark S, Wegrzyn J, Hwang S-R. Proteases for processing

444   proneuropeptides into peptide neurotransmitters and hormones. Annu Rev Pharmacol Toxicol. 2008;48:

445   393–423. doi:10.1146/annurev.pharmtox.48.113006.094812

446   3.   Hökfelt T, Broberger C, Xu ZQ, Sergeyev V, Ubink R, Diez M. Neuropeptides--an overview.

447   Neuropharmacology. 2000;39: 1337–56. doi:https://doi.org/10.1016/S0028-3908(00)00010-1

448   4.   Emeson RB. Hypothalamic peptidyl-glycine alpha-amidating monooxygenase: preliminary

449   characterization. J Neurosci. 1984;4: 2604–13. doi:https://doi.org/10.1523/JNEUROSCI.04-10-
450   02604.1984

451   5.   Harada M, Fukasawa KM, Fukasawa K, Nagatsu T. Inhibitory action of proline-containing peptides on
452   Xaa-Pro-dipeptidylaminopeptidase. Biochim Biophys Acta. 1982;705: 288–90.
453   doi:https://doi.org/10.1016/0167-4838(82)90191-1

454   6.   Conlon JM, Larhammar D. The evolution of neuroendocrine peptides. Gen Comp Endocrinol. 2005;142:
455   53–9. doi:10.1016/j.ygcen.2004.11.016

456   7.   Baggerman G, Cerstiaens A, De Loof A, Schoofs L. Peptidomics of the larval Drosophila melanogaster
457   central nervous system. J Biol Chem. 2002;277: 40368–74. doi:10.1074/jbc.M206257200

458   8.   Clynen E, Baggerman G, Veelaert D, Cerstiaens A, Van der Horst D, Harthoorn L, et al. Peptidomics of
459   the pars intercerebralis-corpus cardiacum complex of the migratory locust, Locusta migratoria. Eur J
460   Biochem. 2001;268: 1929–39. doi:https://doi.org/10.1046/j.1432-1327.2001.02067.x

461   9.   Fricker LD, Lim J, Pan H, Che F-Y. Peptidomics: identification and quantification of endogenous
462   peptides in neuroendocrine tissues. Mass Spectrom Rev. 2006;25: 327–44. doi:10.1002/mas.20079

463   10.   De Haes W, Van Sinay E, Detienne G, Temmerman L, Schoofs L, Boonen K. Functional
464   neuropeptidomics in invertebrates. Biochim Biophys Acta. 2014; doi:10.1016/j.bbapap.2014.12.011

465   11.   Dallas DC, Guerrero A, Parker EA, Robinson RC, Gan J, German JB, et al. Current peptidomics:
466   Applications, purification, identification, quantification and functional analysis. Proteomics. 2014;
467   doi:10.1002/pmic.201400310

468   12.   Hayakawa E, Menschaert G, De Bock P-J, Luyten W, Gevaert K, Baggerman G, et al. Improving the
469   identification rate of endogenous peptides using electron transfer dissociation and collision-induced
470   dissociation. J Proteome Res. 2013;12: 5410–21. doi:10.1021/pr400446z

471   13.   Hummon AB, Richmond TA, Verleyen P, Baggerman G, Huybrechts J, Ewing MA, et al. From the
472   genome to the proteome: uncovering peptides in the Apis brain. Science. 2006;314: 647–9.
473   doi:10.1126/science.1124128

474   14.   Christie AE. Prediction of the peptidomes of Tigriopus californicus and Lepeophtheirus salmonis
475   (Copepoda, Crustacea). Gen Comp Endocrinol. 2014;201: 87–106. doi:10.1016/j.ygcen.2014.02.015

476   15.   Hauser F, Neupert S, Williamson M, Predel R, Tanaka Y, Grimmelikhuijzen CJP. Genomics and

477   peptidomics of neuropeptides and protein hormones present in the parasitic wasp Nasonia vitripennis. J

478   Proteome Res. 2010;9: 5296–310. doi:10.1021/pr100570j

479 16. Dircksen H, Neupert S, Predel R, Verleyen P, Huybrechts J, Strauss J, et al. Genomics, transcriptomics,

480   and peptidomics of Daphnia pulex neuropeptides and protein hormones. J Proteome Res. 2011;10: 4478–

481   504. doi:10.1021/pr200284e

482 17. Jékely G. Global view of the evolution and diversity of metazoan neuropeptide signaling. Proc Natl

483   Acad Sci U S A. 2013;110: 8702–7. doi:10.1073/pnas.1221833110

484 18. De Loof A, Schoofs L. Homologies between the amino acid sequences of some vertebrate peptide

485   hormones and peptides isolated from invertebrate sources. Comp Biochem Physiol B. 1990;95: 459–68.

486   doi:https://doi.org/10.1016/0305-0491(90)90003-C

487 19. Watanabe H, Fujisawa T, Holstein TW. Cnidarians and the evolutionary origin of the nervous system.

488   Dev Growth Differ. 2009;51: 167–83. doi:10.1111/j.1440-169X.2009.01103.x

489 20. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, et al. Sea anemone genome

490   reveals ancestral eumetazoan gene repertoire and genomic organization. Science. 2007;317: 86–94.

491   doi:10.1126/science.1139158

492 21. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from

493   transmembrane regions. Nat Methods. 2011;8: 785–6. doi:10.1038/nmeth.1701

494 22. Lengfeld T, Watanabe H, Simakov O, Lindgens D, Gee L, Law L, et al. Multiple Wnts are involved in

495   Hydra organizer formation and regeneration. Dev Biol. 2009;330: 186–99.

496   doi:10.1016/j.ydbio.2009.02.004

497 23. Kusserow A, Pang K, Sturm C, Hrouda M, Lentfer J, Schmidt HA, et al. Unexpected complexity of the

498   Wnt gene family in a sea anemone. Nature. 2005;433: 156–60. doi:10.1038/nature03158

499 24. Grimmelikhuijzen CJ, Rinehart KL, Spencer AN. Isolation of the neuropeptide less than Glu-Trp-Leu-

500   Lys-Gly-Arg-Phe-NH2 (Pol-RFamide II) from the hydromedusa Polyorchis penicillatus. Biochem

501   Biophys Res Commun. 1992;183: 375–82. doi:https://doi.org/10.1016/0006-291X(92)90491-3

502 25. Darmer D, Schmutzler C, Diekhoff D, Grimmelikhuijzen CJ. Primary structure of the precursor for the

503   sea anemone neuropeptide Antho-RFamide (less than Glu-Gly-Arg-Phe-NH2). Proc Natl Acad Sci U S

504   A. 1991;88: 2555–9. doi:https://doi.org/10.1073/pnas.88.6.2555

505   26.   Grimmelikhuijzen CJ, Graff D. Isolation of pyroGlu-Gly-Arg-Phe-NH2 (Antho-RFamide), a

506   neuropeptide from sea anemones. Proc Natl Acad Sci U S A. 1986;83: 9817–21.

507   doi:10.1073/pnas.83.24.9817

508   27.   Leviev I, Grimmelikhuijzen CJ. Molecular cloning of a preprohormone from sea anemones containing

509   numerous copies of a metamorphosis-inducing neuropeptide: a likely role for dipeptidyl aminopeptidase

510   in neuropeptide precursor processing. Proc Natl Acad Sci U S A. 1995;92: 11647–51.

511   doi:https://doi.org/10.1073/pnas.92.25.11647

512   28.   Moosler A, Rinehart KL, Grimmelikhuijzen CJ. Isolation of four novel neuropeptides, the hydra-

513   RFamides I-IV, from Hydra magnipapillata. Biochem Biophys Res Commun. 1996;229: 596–602.

514   doi:10.1006/bbrc.1996.1849

515   29.   Takahashi T, Muneoka Y, Lohmann J, Lopez de Haro MS, Solleder G, Bosch TC, et al. Systematic

516   isolation of peptide signal molecules regulating development in hydra: LWamide and PW families. Proc

517   Natl Acad Sci U S A. 1997;94: 1241–6. doi:https://doi.org/10.1073/pnas.94.4.1241

518   30.   Darmer D, Hauser F, Nothacker HP, Bosch TC, Williamson M, Grimmelikhuijzen CJ. Three different

519   prohormones yield a variety of Hydra-RFamide (Arg-Phe-NH2) neuropeptides in Hydra magnipapillata.

520   Biochem J. 1998;332 (Pt 2: 403–12. doi:DOI: 10.1042/bj3320403

521   31.   Hayakawa E, Takahashi T, Nishimiya-Fujisawa C, Fujisawa T. A novel neuropeptide (FRamide) family

522   identified by a peptidomic approach in Hydra magnipapillata. FEBS J. 2007;274: 5438–48.

523   doi:10.1111/j.1742-4658.2007.06071.x

524   32.   Gajewski M, Leitz T, Schloßherr J, Plickert G. LWamides from Cnidaria constitute a novel family of

525   neuropeptides with morphogenetic activity. Roux's Arch Dev Biol. 1996;205: 232–242.

526   doi:10.1007/BF00365801

527   33.   Takahashi T, Koizumi O, Ariura Y, Romanovitch A, Bosch TC, Kobayakawa Y, et al. A novel

528   neuropeptide, Hym-355, positively regulates neuron differentiation in Hydra. Development. 2000;127:

529   997–1005. doi:DOI: 10.1016/S1095-6433(99)90367-7

530   34.   Fälth M, Sköld K, Svensson M, Nilsson A, Fenyö D, Andren PE. Neuropeptidomics strategies for

531   specific and sensitive identification of endogenous peptides. Mol Cell Proteomics. 2007;6: 1188–97.

532   doi:10.1074/mcp.M700016-MCP200

533   35.   Yum S, Takahashi T, Hatta M, Fujisawa T. The structure and expression of a preprohormone of a

534   neuropeptide, Hym-176 in Hydra magnipapillata. FEBS Lett. 1998;439: 31–4.

535   doi:https://doi.org/10.1016/S0014-5793(98)01314-3

536   36.   Schmutzler C, Darmer D, Diekhoff D, Grimmelikhuijzen CJ. Identification of a novel type of processing

537   sites in the precursor for the sea anemone neuropeptide Antho-RFamide (<Glu-Gly-Arg-Phe-NH2) from

538   Anthopleura elegantissima. J Biol Chem. 1992;267: 22534–41. Available:

539   http://www.ncbi.nlm.nih.gov/pubmed/1429603

540   37.   Reinscheid RK, Grimmelikhuijzen CJ. Primary structure of the precursor for the anthozoan neuropeptide

541   antho-RFamide from Renilla köllikeri: evidence for unusual processing enzymes. J Neurochem.

542   1994;62: 1214–22. doi:https://doi.org/10.1046/j.1471-4159.1994.62031214.x

543   38.   Schoofs L, Holman GM, Hayes TK, Kochansky JP, Nachman RJ, De Loof A. Locustatachykinin III and

544   IV: two additional insect neuropeptides with homology to peptides of the vertebrate tachykinin family.

545   Regul Pept. 1990;31: 199–212. doi:https://doi.org/10.1016/0167-0115(90)90006-I

546   39.   Lundquist CT, Clottens FL, Holman GM, Nichols R, Nachman RJ, Nässel DR. Callitachykinin I and II,

547   two novel myotropic peptides isolated from the blowfly, Calliphora vomitoria, that have resemblances to

548   tachykinins. Peptides. 1994;15: 761–8. doi:http://dx.doi.org/10.1016/0196-9781(94)90027-2

549   40.   Neupert S, Russell WK, Russell DH, López JD, Predel R, Nachman RJ. Neuropeptides in Heteroptera:

550   identification of allatotropin-related peptide and tachykinin-related peptides using MALDI-TOF mass

551   spectrometry. Peptides. 2009;30: 483–8. doi:10.1016/j.peptides.2008.11.009

552   41.   Anastasi A, Erspamer V. The isolation and amino acid sequence of eledoisin, the active endecapeptide of

553   the posterior salivary glands of Eledone. Arch Biochem Biophys. 1963;101: 56–65.

554   doi:https://doi.org/10.1016/0003-9861(63)90533-2

555   42.   Chiwakata C, Brackmann B, Hunt N, Davidoff M, Schulze W, Ivell R. Tachykinin (substance-P) gene

556   expression in Leydig cells of the human and mouse testis. Endocrinology. 1991;128: 2441–8.

557   doi:10.1210/endo-128-5-2441

558   43.   Veenstra JA. Neuropeptide evolution: neurohormones and neuropeptides predicted from the genomes of

559   Capitella teleta and Helobdella robusta. Gen Comp Endocrinol. 2011;171: 160–75.

560   doi:10.1016/j.ygcen.2011.01.005

561   44.   Sato Y, Oguchi M, Menjo N, Imai K, Saito H, Ikeda M, et al. Precursor polyprotein for multiple

562   neuropeptides secreted from the suboesophageal ganglion of the silkworm Bombyx mori:

563   characterization of the cDNA encoding the diapause hormone precursor and identification of additional

564   peptides. Proc Natl Acad Sci U S A. 1993;90: 3251–5. doi:https://doi.org/10.1073/pnas.90.8.3251

565   45.   Duportets L, Gadenne C, Couillaud F. A cDNA, from Agrotis ipsilon, that encodes the pheromone

566   biosynthesis activating neuropeptide (PBAN) and other FXPRL peptides. Peptides. 1999;20: 899–905.

567   doi:https://doi.org/10.1016/S0196-9781(99)00079-0

568   46.   Hauser F, Williamson M, Cazzamali G, Grimmelikhuijzen CJP. Identifying neuropeptide and protein

569   hormone receptors in Drosophila melanogaster by exploiting genomic data. Brief Funct Genomic

570   Proteomic. 2006;4: 321–30. doi:10.1093/bfgp/eli003

571   47.   Mori K, Miyazato M, Ida T, Murakami N, Serino R, Ueta Y, et al. Identification of neuromedin S and its

572   possible role in the mammalian circadian oscillator system. EMBO J. 2005;24: 325–35.

573   doi:10.1038/sj.emboj.7600526

574   48.   Brighton PJ, Szekeres PG, Willars GB. Neuromedin U and its receptors: structure, function, and

575   physiological roles. Pharmacol Rev. 2004;56: 231–48. doi:10.1124/pr.56.2.3

576   49.   Schaffer MH, Noyes BE, Slaughter CA, Thorne GC, Gaskell SJ. The fruitfly Drosophila melanogaster

577   contains a novel charged adipokinetic-hormone-family peptide. Biochem J. 1990;269: 315–20. doi:DOI:

578   10.1042/bj2690315

579   50.   Ma M, Bors EK, Dickinson ES, Kwiatkowski MA, Sousa GL, Henry RP, et al. Characterization of the

580   Carcinus maenas neuropeptidome by mass spectrometry and functional genomics. Gen Comp

581   Endocrinol. 2009;161: 320–34. doi:10.1016/j.ygcen.2009.01.015

582   51.   Gäde G, Rinehart KL. Primary structures of hypertrehalosaemic neuropeptides isolated from the corpora

583   cardiaca of the cockroaches Leucophaea maderae, Gromphadorhina portentosa, Blattella germanica and

584   Blatta orientalis and of the stick insect Extatosoma tiaratum assigned by t. Biol Chem Hoppe Seyler.

585   1990;371: 345–54.

586   52.   Zhang L, Tello JA, Zhang W, Tsai P-S. Molecular cloning, expression pattern, and immunocytochemical

587   localization of a gonadotropin-releasing hormone-like molecule in the gastropod mollusk, Aplysia

588   californica. Gen Comp Endocrinol. 2008;156: 201–9. doi:10.1016/j.ygcen.2007.11.015

589   53.   Kawada T, Ogasawara M, Sekiguchi T, Aoyama M, Hotta K, Oka K, et al. Peptidomic analysis of the

590   central nervous system of the protochordate, Ciona intestinalis: homologs and prototypes of vertebrate

591   peptides and novel peptides. Endocrinology. 2011;152: 2416–27. doi:10.1210/en.2010-1348

592   54.   Mason AJ, Hayflick JS, Zoeller RT, Young WS, Phillips HS, Nikolics K, et al. A deletion truncating the

593   gonadotropin-releasing hormone gene is responsible for hypogonadism in the hpg mouse. Science.

594   1986;234: 1366–71. doi:DOI: 10.1126/science.3024317

595   55.   Predel R, Nachman RJ, Gäde G. Myostimulatory neuropeptides in cockroaches: structures, distribution,

596   pharmacological activities, and mimetic analogs. J Insect Physiol. 2001;47: 311–24.

597   doi:https://doi.org/10.1016/S0022-1910(00)00129-3

598   56.   Clynen E, Schoofs L. Peptidomic survey of the locust neuroendocrine system. Insect Biochem Mol Biol.

599   2009;39: 491–507. doi:10.1016/j.ibmb.2009.06.001

600   57.   Schoofs L, De Loof A, Van Hiel MB. Neuropeptides as Regulators of Behavior in Insects. Annu Rev

601   Entomol. 2017;62: 35–52. doi:10.1146/annurev-ento-031616-035500

602   58.   Park Y, Kim Y-J, Adams ME. Identification of G protein-coupled receptors for Drosophila PRXamide

603   peptides, CCAP, corazonin, and AKH supports a theory of ligand-receptor coevolution. Proc Natl Acad

604   Sci U S A. 2002;99: 11423–8. doi:10.1073/pnas.162276199

605   59.   Lindemans M, Janssen T, Beets I, Temmerman L, Meelkop E, Schoofs L. Gonadotropin-releasing

606   hormone and adipokinetic hormone signaling systems share a common evolutionary origin. Front

607   Endocrinol (Lausanne). 2011;2: 16. doi:10.3389/fendo.2011.00016

608   60.   Marlow HQ, Srivastava M, Matus DQ, Rokhsar D, Martindale MQ. Anatomy and development of the

609   nervous system of Nematostella vectensis, an anthozoan cnidarian. Dev Neurobiol. 2009;69: 235–54.

610   doi:10.1002/dneu.20698

611   61.   Koizumi O. Nerve ring of the hypostome in hydra: is it an origin of the central nervous system of

612   bilaterian animals? Brain Behav Evol. 2007;69: 151–9. doi:10.1159/000095204

613   62.   Watanabe H, Kuhn A, Fushiki M, Agata K, Özbek S, Fujisawa T, et al. Sequential actions of β-catenin

614   and Bmp pattern the oral nerve net in Nematostella vectensis. Nat Commun. 2014;5: 5536.

615   doi:10.1038/ncomms6536

616   63.   Petruzziello F, Fouillen L, Wadensten H, Kretz R, Andren PE, Rainer G, et al. Extensive characterization

617   of Tupaia belangeri neuropeptidome using an integrated mass spectrometric approach. J Proteome Res.

618   2012;11: 886–96. doi:10.1021/pr200709j

619   64.   Zhang X, Petruzziello F, Zani F, Fouillen L, Andren PE, Solinas G, et al. High identification rates of

620   endogenous neuropeptides from mouse brain. J Proteome Res. 2012;11: 2819–27.

621   doi:10.1021/pr3001699

622   65.   Sasaki K, Osaki T, Minamino N. Large-scale identification of endogenous secretory peptides using

623   electron transfer dissociation mass spectrometry. Mol Cell Proteomics. 2013;12: 700–9.

624   doi:10.1074/mcp.M112.017400

625   66.   Southey BR, Amare A, Zimmerman T a, Rodriguez-Zas SL, Sweedler J V. NeuroPred: a tool to predict

626   cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides. Nucleic Acids

627   Res. 2006;34: W267-72. doi:10.1093/nar/gkl161

628   67.   Darmer D, Hauser F, Nothacker HP, Bosch TC, Williamson M, Grimmelikhuijzen CJ. Three different

629   prohormones yield a variety of Hydra-RFamide (Arg-Phe-NH2) neuropeptides in Hydra magnipapillata.

630   Biochem J. 1998;332 (Pt 2: 403–12. doi:DOI: 10.1042/bj3320403

631   68.   Leung PS, Shaw C, Maule AG, Thim L, Johnston CF, Irvine GB. The primary structure of neuropeptide

632   F (NPF) from the garden snail, Helix aspersa. Regul Pept. 1992;41: 71–81.

633   doi:https://doi.org/10.1016/0167-0115(92)90515-V

634   69.   Brown MR, Crim JW, Arata RC, Cai HN, Chun C, Shen P. Identification of a Drosophila brain-gut

635   peptide related to the neuropeptide Y family. Peptides. 1999;20: 1035–42.

636   doi:https://doi.org/10.1016/S0196-9781(99)00097-2

637   70.   Cupo A, Proux J. Biochemical characterization of a vasopressin-like neuropeptide in Locusta migratoria.

638   Evidence of high molecular weight protein encoding vasopressin sequence. Neuropeptides. 1983;3: 309–

639   18. doi:https://doi.org/10.1016/0143-4179(83)90048-3

640   71.   Cruz LJ, de Santos V, Zafaralla GC, Ramilo CA, Zeikus R, Gray WR, et al. Invertebrate

641   vasopressin/oxytocin homologs. Characterization of peptides from Conus geographus and Conus straitus

642   venoms. J Biol Chem. 1987;262: 15821–4.

643   72.   Schoofs L, Danger JM, Jegou S, Pelletier G, Huybrechts R, Vaudry H, et al. NPY-like peptides occur in

644   the nervous system and midgut of the migratory locust, Locusta migratoria and in the brain of the grey

645   fleshfly, Sarcophaga bullata. Peptides. 9: 1027–36. doi:https://doi.org/10.1016/0196-9781(88)90084-8

646 73. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, et al. Sea anemone genome

647   reveals ancestral eumetazoan gene repertoire and genomic organization. Science. 2007;317: 86–94.

648   doi:10.1126/science.1139158

649 74. Anctil M. Chemical transmission in the sea anemone Nematostella vectensis: A genomic perspective.

650   Comp Biochem Physiol Part D Genomics Proteomics. Elsevier Inc.; 2009;4: 268–89.

651   doi:10.1016/j.cbd.2009.07.001

652 75. Jékely G. Global view of the evolution and diversity of metazoan neuropeptide signaling. Proc Natl

653   Acad Sci U S A. 2013;110: 8702–7. doi:10.1073/pnas.1221833110

654 76. Mirabeau O, Joly J-S. Molecular evolution of peptidergic signaling systems in bilaterians. Proc Natl

655   Acad Sci U S A. 2013;110: E2028-37. doi:10.1073/pnas.1219956110

656

657

658

Supporting Information

660   **S1 Table.** The structures of the cnidarian neuropeptides.

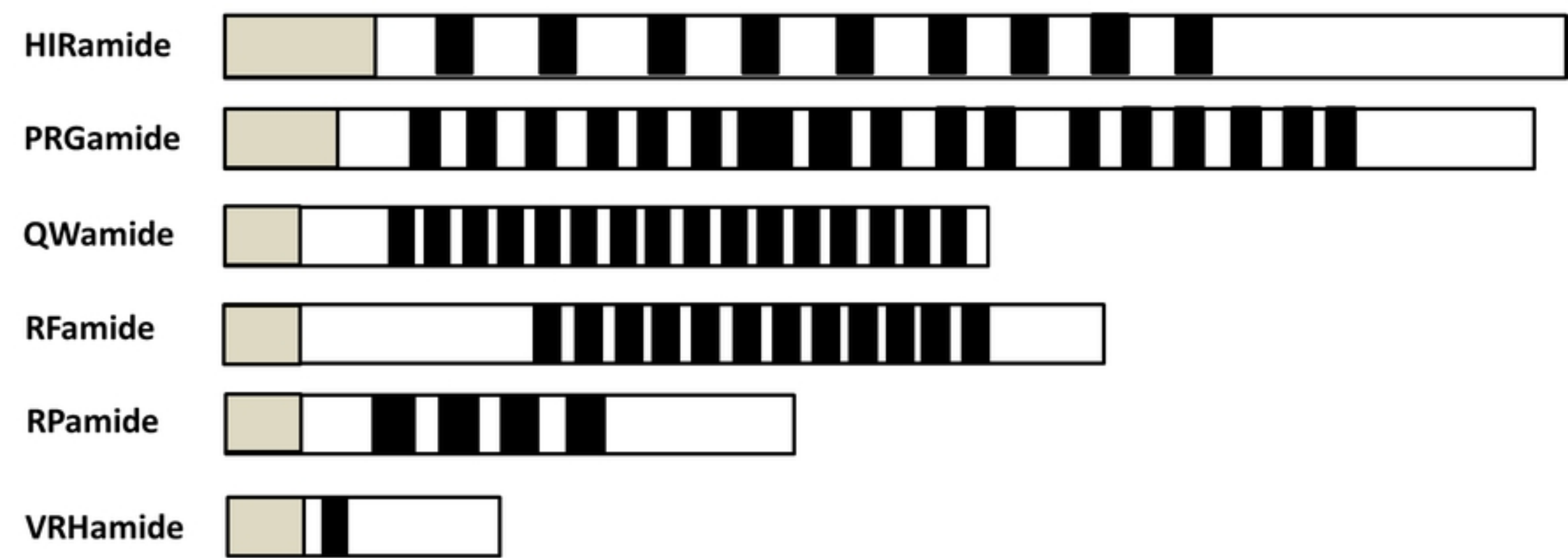661   **S2 Table.** Details of the neuropeptide precursor genes.

662

Fig 1

Fig 2

HIRamide

PRGamide

QWamide

RFamide

RPamide

VRHamide

Fig 3

A
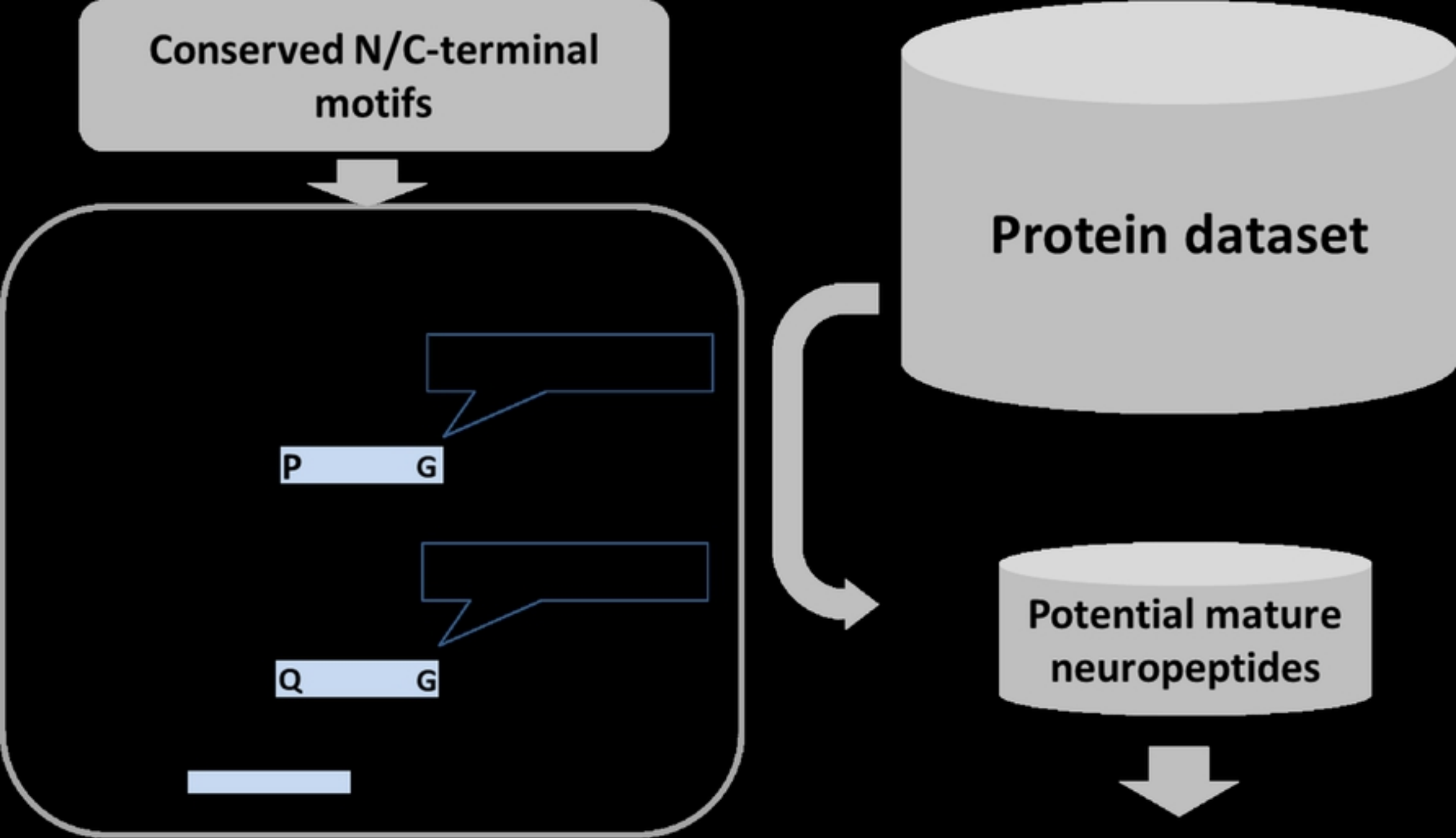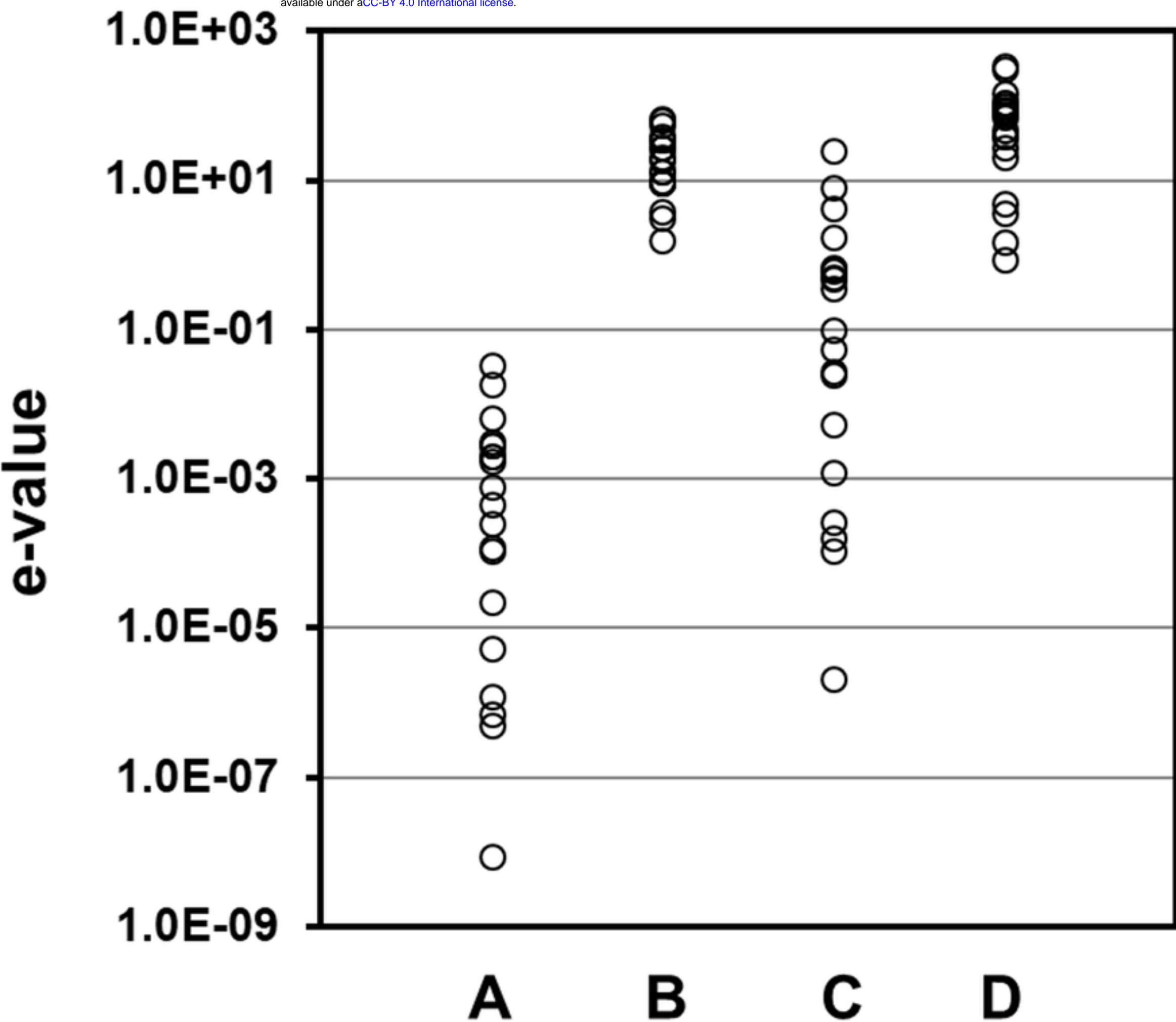
QPPIDL**SPAA**Y**F**H**IR**	HIRamide (*Nematostella vectensis*)
QPPMIDL**SE**PA**F**F**H**I**R**	HIRamide (*Nematostella vectensis*)
QPPYLDL**GE**P**S**F**F**H**I**R**	HIRamide (*Nematostella vectensis*)
QPPYLDL**T**P**S**Y**F**H**I**R**	HIRamide (*Nematostella vectensis*)
A**P**T**A**F**YGV**R	Callitachykinin-1 (*Calliphora vomitoria*)
AP**SS**N**S**F**MGM**R	Tachykinin-related peptide 3 (*Daphnia pulex*)
APA**A**G**F**F**GM**R	Tachykinin-related peptide 2 (*Acrosternum hilare*)
F**S**PK**S**F**H**F**S**R	Tacykinin (*Helobdella robusta*)
HK**T**D**S**F**VGLM	Neurokinin A (*Mus musculus*)
RPK**P**QQ**F**FGLM	Substance P (*Mus musculus*)

Cnidaria
Arthropoda
Annelida
Chordata

B

GPRGG**R**_ATE**F**GPR**G**	PRGamide (*Nematostella vectensis*)
GPRGG**R**EVN**L**E**G**PR**G**	PRGamide (*Nematostella vectensis*)
TDMKDESDRGAHSE**R**GALW**F**GPR**L**	Diapause hormone (*Bombyx mori*)
DVKDGGADRGAHSD**R**GGMW**F**GPR**I**	Diapause hormone homolog (Potential) (*Agrotis ipsilon*)
TGPSASSGLW**F**GPR**L**	Pyrokinin-1 (*Drosophila melanogaster*)
. . . TVDFPKKDPTTSLG**R**PFF**L**F**R**P**R**N	Neuromedin-S (*Mus musculus*)
FKAEYQSPSVGQSKGYF**L**F**R**P**R**N	Neuromedin-U (*Mus musculus*)

Cnidaria
Arthropoda
Chordata

C

IPP**Q**GFR**FS**_**Q**W	QWamide (*Nematostella vectensis*)
IPP**Q**G**L**R**FN**_**Q**W	QWamide (*Nematostella vectensis*)
QC**Q**_**L**T**FS**PD**W**	Adipokinetic hormone (*Drosophila melanogaster*)
Q**L**N**FS**PG**W**	Red pigment-concentrating hormone (*Carcinus maenas*)
Q_VN**FS**PN**W**	Hypertrehalosaemic factor 1 (*Blatta orientalis*)
Q_VN**FS**PG**W**GT	Hypertrehalosaemic hormone (*Rhyparobia maderae*)
QNYH**FS**NG**W**YAG	Gonadotropin-releasing hormone (*Aplysia californica*)
QHW**S**KGYSPG	t-GnRH-3 (*Ciona intestinalis*)
QHW**S**YGLRPG	Gonadoliberin-1 (*Mus musculus*)

Cnidaria
Arthropoda
Mollusca
Chordata

Fig 4

| HIRamide | PRGamide | RFamide | VRHamide |

Fig 5