

# A “Love” Dart Allo hormone Identified in the Mucous Glands of Hermaphroditic Land Snails<sup>\*[5]</sup>

Received for publication, November 22, 2015, and in revised form, January 14, 2016. Published, JBC Papers in Press, January 27, 2016, DOI 10.1074/jbc.M115.704395

Michael J. Stewart<sup>‡</sup>, Tianfang Wang<sup>‡</sup>, Joris M. Koene<sup>§</sup>, Kenneth B. Storey<sup>¶</sup>, and Scott F. Cummins<sup>‡1</sup>

From the <sup>‡</sup>Genecology Research Centre, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Maroochydore, Queensland 4558, Australia, the <sup>§</sup>Department of Ecological Science, Faculty of Earth and Life Sciences, Vrije Universiteit, 1081HV Amsterdam, The Netherlands, and the <sup>¶</sup>Institute of Biochemistry and Department of Biology, Carleton University, Ottawa, Ontario K1S 5B6, Canada

Animals have evolved many ways to enhance their own reproductive success. One bizarre sexual ritual is the “love” dart shooting of helioid snails, which has courted many theories regarding its precise function. Acting as a hypodermic needle, the dart transfers an allo hormone that increases paternity success. Its precise physiological mechanism of action within the recipient snail is to close off the entrance to the sperm digestion organ via a contraction of the copulatory canal, thereby delaying the digestion of most donated sperm. In this study, we used the common garden snail *Cornu aspersum* to identify the allo hormone that is responsible for this physiological change in the female system of this simultaneous hermaphrodite. The love dart allo hormone (LDA) was isolated from extracts derived from mucous glands that coat the dart before it is stabbed through the partner’s body wall. We isolated LDA from extracts using bioassay-guided contractility measurement of the copulatory canal. LDA is encoded within a 235-amino acid precursor protein containing multiple cleavage sites that, when cleaved, releases multiple bioactive peptides. Synthetic LDA also stimulated copulatory canal contractility. Combined with our finding that the protein amino acid sequence resembles previously described molluscan buccalin precursors, this indicates that LDA is partially conserved in helioid snails and less in other molluscan species. In summary, our study provides the full identification of an allo hormone that is hypodermically injected via a love dart. More importantly, our findings have important consequences for understanding reproductive biology and the evolution of alternative reproductive strategies.

In animal species with separate sexes, mate choice and precopulatory sexual selection of females are influenced by their reproductive physiology and behavior (1), whereas competition between males to generate offspring is generally associated with morphological and physical characteristics (1–3). However, such competition for fertilization can continue after copula-

tion, at the level of the sperm, and this process seems to have become an especially important evolutionary driving force among a group of species with a different reproductive strategy: simultaneous hermaphrodites that do not self-fertilize (4–6).

Helioid land snail copulation lasts 2–6 h and includes the unique use of calcareous (calcium carbonate) “love” darts that are pierced through the body wall of the mating partner during courtship (7–10). The mating strategy often requires a precopulatory courtship with associated behaviors. In the snail *Cornu aspersum* (formerly known as *Cantareus aspersus* and *Helix aspersa*), there are three major components, described as introductory behavior (reciprocal tactile and oral contacts), calcareous love dart shooting, and copulation (11). Meanwhile, the mating behaviors for the snail *Helix pomatia* appear to be quite different and require elaborate caressing with the tentacles and oral lappets accompanied by intertwining and body contractions (12). However, this is still followed by the injection of a calcareous love dart and copulation.

Research has implicated the mucus that coats these calcareous love darts as containing compounds that enhance paternity success; thus the love dart is used as a tool for increasing fertilization success of the shooter (10, 13–16). Love darts are characteristic features of the reproductive tract for the hermaphroditic land snail families of Helicidae and Ariophantidae (12). These darts are assembled within a thick-walled dart sac and have now been described in numerous species, showing distinct species-specific structural architecture (17). A mucous gland is in close proximity to the dart sac, so that darts can be coated with mucus prior to release. The exact function of the mucus-covered dart remained a mystery for many years, although several explanations had been suggested. One such explanation that gained prominence was that the love dart existed as a nuptial gift of calcium for the mating partner, to be used for egg production (4, 14). This was later disproved, because the amount of calcium in one dart is roughly equal to that of one egg, and thus it would not contribute significantly to an average clutch of 59 eggs (for *C. aspersum*) (18).

More recent research has provided better insight, showing that when a mucus-covered dart is successfully stabbed into the partner, there is a corresponding increase in paternity success (13). Without mating with a mucus-covered dart, the spermatophore placed into the spermatophore-receiving organ of their partner is almost entirely digested, leaving only a small number (0.1%) of the spermatozoa to survive and escape to the more advanced regions of the female reproductive system for

\* This work was supported by Grains Research and Development Corporation Innovation and Investment Grants USU00001 (to M. J. S., T. W., and S. F. C.) and by funds from the Australian Research Council (to S. F. C. and K. B. S.). The authors declare that they have no conflicts of interest with the contents of this article.

[5] This article contains supplemental Tables S1 and S2, Fig. S1, and Files S1 and S2.

<sup>1</sup> To whom correspondence should be addressed: Dept. of Science and Education, University of the Sunshine Coast, 90 Sippy Downs Dr., Sippy Downs, QLD 4556, Australia. Tel.: 61-54565501; E-mail: scummins@usc.edu.au.

storage until fertilization occurs (19). The physiological mechanism of action seems to be through reconfiguration of the spermatophore-receiving organ copulatory canal via peristaltic movements, which in turn increases the number of escaping spermatozoa (13). Because the love darts alone do not increase paternity success, the mucus is implicated as a carrier of the active substance(s) (10).

Dart shooting probably evolved as the result of sperm competition, so that snails can optimize their reproductive success. Thus despite arguments against it, sexual selection does play an important role in hermaphrodite evolution through sexually antagonistic selection: darting benefits male function by increasing paternity (17). The active agent in the gland mucus fulfills the requirements for the substance to be classified as an allohormone; it enters the circulatory system of the conspecific via inoculation by the dart and stimulates physiological and behavioral changes (20). In this study, we analyzed the contents of the mucous glands of *C. aspersum* and isolated the love dart allohormone (LDA)<sup>2</sup> that stimulates contraction of the copulatory canal. We further confirm the bioactivity using a synthetic LDA. We show that the LDA precursor is found in abundance within the mucous glands of *C. aspersum*, as well as another helicid, *Theba pisana*.

## Experimental Procedures

**Snails**—All use of animals for this research was carried out in accordance with the recommendations set by the Animal Ethics Committee, University of the Sunshine Coast. Adult garden snails (*C. aspersum*) were purchased from commercial culturists (Glasshouse Gourmet Snails), transported, and stored at University of the Sunshine Coast helicid facilities during the months of April and May in 2013 and 2014. *T. pisana* were obtained from Warooka, South Australia, and the grains belt of South Yorke Peninsula, South Australia, and transported to the University of the Sunshine Coast helicid facilities. All snails were housed in moist nylon mesh-covered pens on a 12-h light/dark photoperiod and fed alternating days with slices of cucumber or carrot until used for bioassay.

**In Vitro Assay 1: Crude Mucus Extracts**—The contractile activity of the *C. aspersum* copulatory canal (intersecting T-junction of the bursa tract and bursa tract diverticulum as described by Koene and Chase (13) (see Fig. 1) was monitored after exposure to *C. aspersum* crude mucous gland preparations (extracts taken from 28 mucous glands; mucus proteins separated under 3 kDa; mucus proteins above 3 kDa), 10  $\mu$ M 5-HT (positive control), or 10  $\mu$ M BSA using an isometric force transducer (i-FOT Rev 3.0; GlobalTown Microtech, Sarasota, FL) linked to a compact unichannel bridge amplifier (BRAM-4B; GlobalTown) and a polygraph monitoring system (PowerLabs<sup>TM</sup> connected to Chart 5; ADI Instruments). The addition of treatments was randomized to avoid any order effects of testing (*i.e.* added effects, false contractions, and instances of tissue tiredness) and delivered approximately every 90 s to the copulatory canal suspended between the transducer and a fixed

0.1-mm hook using nylon ligatures, in 20- $\mu$ l aliquots to a 15-ml organ bath containing oxygenated (95% O<sub>2</sub>, 5% CO<sub>2</sub>), filtered, molluscan Ringer solution (21) at 25 °C. The effects of each test substance were examined for the height of contraction amplitude and duration. In all assays, statistical analysis was performed using Pearson's chi-square test or Fisher's exact test as appropriate. In all assays, the *p* value can be deemed at the *p* = 0.05 level or lower.

**RP-HPLC of <3-kDa Extracts and Preparation for in Vitro Assay**—Paired mucous glands were removed from 14 specimens of *C. aspersum* and placed into separate glass convex dishes containing 1 ml of chilled MilliQ water treated with 0.1% TFA. From the pools of mucous glands, their tips were cut with a scalpel blade and with fine forceps, milked by dragging the forceps over the length of the mucous glands. Released mucous was then collected and dispensed into a 15-ml Falcon tube and washed in 10 ml of 0.1% TFA. Soluble compounds were then extracted through sonication. The mucous extracts were then centrifuged at 7,500  $\times$  *g* for 10 min at 4 °C, and the supernatant was acidified with 0.1% TFA. Following this, the supernatant was passed through a C18 Sep-Pack Vac cartridge (5 g; Waters, Rydalmere, Australia) pretreated with 5 ml of 100% CH<sub>3</sub>CN, followed by 10 ml of MilliQ water treated with 0.1% TFA. The biomolecules were then eluted with 60% CH<sub>3</sub>CN, 0.1% TFA. Prior to RP-HPLC, samples were lyophilized and then resuspended in 1 ml of single-distilled H<sub>2</sub>O. If required to separate out proteins based on molecular weight, the semipurified mucous was then passed through a 2-ml Ultracel (Amicon Ultra; Merck) on a swing bucket centrifuge at 4,000  $\times$  *g*. Separated proteins under 3 kDa were lyophilized for RP-HPLC, whereas all proteins over 3 kDa were further purified via C18 Sep-Pack cartridges and prepared for RP-HPLC as above. All samples aside from proteins over 3 kDa were run through RP-HPLC on a PerkinElmer Life Sciences series 200 pump/autosampler, Flexar photodiode array detector, and Chromera v3.2 software, using an Agilent Zorbax 300 SB-C18 column (internal diameter 4.8 mm  $\times$  150 mm and particle size of 5  $\mu$ m) for 40 min with a linear gradient of 0–60% CH<sub>3</sub>CN containing 0.1% TFA. Extracts that were not separated into <3 kDa and >3 kDa prior to RP-HPLC were collected every min for 5 min as pooled fractions and detected at wavelengths of 210 and 280 nm. This process was also repeated to collect 1-min individual fractions. In both occasions, fractions (pooled or individual) were then lyophilized and resuspended in 80  $\mu$ l of single-distilled H<sub>2</sub>O for use in bioassay. Half of the sample was then used for *in vitro* assays (see below), and the other half was acidified with 0.1% TFA and desalted using ziptips (Waters) prior to proteome and MALDI-TOF analysis. Mucus extracts of <3-kDa fractions were collected every min and processed the same way.

**In Vitro Assay 2: Pooled RP-HPLC Fractions and Individual Fractions 21–25**—The contractile activity of the *C. aspersum* copulatory canal was monitored using conditions as described above (*in vitro* assay of crude mucus extracts). In assay 2a, lyophilized pooled mucous gland fractions (*i.e.* fractions 1–5, fractions 6–10, etc.) corresponding to 5-min intervals were delivered in 20- $\mu$ l aliquots every 90 s to the copulatory canal (suspended between the transducer and a fixed 0.1-mm hook

<sup>2</sup>The abbreviations used are: LDA, love dart allohormone; contig, group of overlapping clones; SEM, scanning electron microscopy; RP, reverse phase; msc, mucus-secreting cell; 5-HT, 5-hydroxytryptamine.

## Allohormone in Helicid Snails

using nylon ligatures) held in a 15-ml organ bath containing oxygenated (95% O<sub>2</sub>, 5% CO<sub>2</sub>), filtered, molluscan Ringer at 25 °C. The effects of each fraction were examined for the height of contraction amplitude and duration. In assay 2b, after identifying the positive pooled mucus fraction set (*viz.*, pooled fractions 21–25), fractions 21–25 were then separated via RP-HPLC (using the preparative conditions described above) with fractions collected in 1-min aliquots. Individual fractions were lyophilized and resuspended in 80 μl of single-distilled H<sub>2</sub>O and delivered in 20-μl aliquots to test their effects on the contractile activity of the *C. aspersum* copulatory canal using the same testing conditions and parameters as described in *in vitro* assay 2a. In both assays, Ringer and 10 μM BSA was used as negative controls, and 10 μM 5-HT was used as a positive control.

**RP-HPLC of Fraction 21**—Mucous glands of *C. aspersum* ( $n = 32$ ) were placed into separate convex glass dishes containing 1 ml of chilled MilliQ water treated with 0.1% TFA. From the pools of mucous glands, their tips were then cut with a scalpel blade and fine forceps and milked. Released mucus was then collected and dispensed in to a 15-ml Falcon tube and made up to 10 ml in 0.1% TFA. Soluble compounds were then extracted through sonication. The mucous extracts were centrifuged at  $7,500 \times g$  for 10 min at 4 °C, and the acidified supernatant was collected and passed through a C18 Sep-Pack Vac cartridge (5 g; Waters) pretreated with 5 ml of 100% CH<sub>3</sub>CN, followed by 10 ml of MilliQ water treated with 0.1% TFA. Biomolecules were then eluted with 60% CH<sub>3</sub>CN, 0.1% TFA. The sample was lyophilized and resuspended in 1 ml of single-distilled H<sub>2</sub>O. Following the parameters set above, the sample was run on RP-HPLC using an Agilent Zorbax 300 SB-C18 column (internal diameter, 4.8 mm  $\times$  150 mm and particle size of 5 μm) with the same linear gradient. When sampling reached the start of fraction 21, the eluate was collected every 4 drops (71.43 μl/every 4.3 s) for a total of 14 individual fractionated eluates.

**In Vitro Assay 3: Fractionation of Fraction 21**—Separated fractions obtained from RP-HPLC were then assessed for their effects on the contractile activity of the *C. aspersum* copulatory canal. Using the same monitoring conditions as set out in *in vitro* assays 1 and 2, lyophilized fractions (*i.e.* fraction 21-1, fraction 21-2, etc.) were delivered every 90 s to the copulatory canal (suspended between the transducer and a fixed 0.1-mm hook using nylon ligatures) in 20-μl aliquots to a 15-ml organ bath containing oxygenated (95% O<sub>2</sub>, 5% CO<sub>2</sub>), filtered, molluscan Ringer at 25 °C. The effects of each fraction were examined for height of contraction amplitude and duration. Those fractions that induced contractility at 100% ( $n = 3$ ) were considered positive. The same controls were tested as per *in vitro* assays 1 and 2.

**Illumina cDNA Library Preparation and Bioinformatics**—Approximately 12 mucous glands were removed from 6 *C. aspersum* that had been fed over a 3-day period. The mucous glands were then pooled washed in Millipore water, homogenized in TRIzol reagent (Invitrogen), and processed following the manufacturer's protocols. The RNA pellets were concentrated using 8 M LiCl RNA followed by a wash with 70% ice-cold ethanol. Purified total RNA was then dissolved with 50 μl of

warmed RNase-free water and combined. The quantity and quality of pooled RNA was assessed using UV spectrophotometry (NanoDrop ND-1000). Approximately 2 μg of total RNA was then used for transcriptome analysis using the Illumina sequencing platform following generation of a cDNA library (BGI, Shenzhen, China). Prior to transcriptome assembly and mapping, filters were implemented to remove low quality reads and adaptor sequence.

To predict gene function and identify putative conserved protein domains to the contig set, we compared translated protein sequences deduced from our contigs to multiple functional domain databases using RPS-BLAST and Blast2GO. First, the entire transcribed sets were compared with the SMART, COG conserved domains database, Protein Family database (Pfam), and CDD databases using RPS-BLAST with no expect value threshold cutoff, but only matches with an expect value less than  $1 \times 10^{-10}$  were considered in further analyses. We also mapped our contigs for gene ontology searching the gene ontology database and Blast2GO platform. A conservative set of contigs was obtained using the bioinformatics suite above, and contigs were run through a six-frame ORF filtering criteria (ExpASY-Translate tool and NCBI ORF finder) to provide a *C. aspersum* mucous gland protein database.

**Peptidome Analysis and Mass Spectrometry-MALDI**—MALDI-MS of mucous proteins under 3 kDa were obtained using a 4700 MALDI-TOF/TOF mass spectrometer (Applied Biosystems) with a mass range from  $m/z$  500 to 3000. This TOF/TOF instrument was equipped with an Nd:YAG laser with 355-nm wavelength of <500-ps pulse and a 200-Hz firing rate. Samples were dissolved in 60% acetonitrile (CH<sub>3</sub>CN), 0.1%TFA, and 0.4 μl were placed onto a MALDI-MS sample plate containing 0.4 μl of matrix solution 2-cyano-3-(4-hydroxyphenyl)acrylic acid, made from mass standards kit from AB Sciex, or sinnapinic acid (3-(4-hydroxy-3,5-dimethoxyphenyl)prop-2-enoic acid). After drying at ambient temperature, the sample plate was blown by nitrogen to remove any dust or fibers off the target and then analyzed immediately. MALDI-TOF/TOF data were acquired in the batch mode, and the *S/N* filter was set at 3. The instrument was calibrated using a standard from the mass standards kit. Final spectra resulted from an accumulation of 2000 shots and were processed by 4000 Series Explorer version 3.5.3 (AB Sciex) without smoothing; baseline subtraction was performed with peak width set to 50.

**LC-MS/MS**—Tryptic peptides were analyzed by LC-MS/MS on a Shimadzu Prominence Nano HPLC (Japan) system coupled to a Triple TOF™ 5600 mass spectrometer (AB Sciex) equipped with a nano electrospray ion source. A 6-μl aliquot of each extract was injected onto a 50-mm  $\times$  300-μm C18 trap column (Agilent Technologies). Tryptic digests from each fraction were desalted on the trap column for 5 min using acidified 0.1% (v/v) formic acid (aqueous) at 30 μl min<sup>-1</sup>. The trap column was then placed in line with the analytical nano HPLC column, a 150-mm  $\times$  75-μm 300SBC18, 3.5 μm (Agilent Technologies) for mass spectrometry analysis. A linear gradient at a flow rate of 300 nl min<sup>-1</sup>, from 1–40% solvent B over 35 min, followed by a steeper gradient from 40% to 80% solvent B in 5 min were utilized for peptide elution. Solvent B was held at 80% for 5 min for washing the column and returned to 1% solvent B

for equilibration prior to the next sample injection. Solvent A was 0.1% (v/v) formic acid (aqueous), and solvent B consisted of 0.1% (v/v) formic acid in 90% (v/v) acetonitrile. The ion spray voltage was set to 2400 V, declustering potential was 100 V, the curtain gas flow was 25, nebulizer gas 1 was 12, and the heated interface was at 150 °C. The mass spectrometer acquired 500-ms full scan TOF-MS data followed by 20 × 50-ms full scan product ion data in an information-dependent acquisition mode. Full scan TOF-MS data were acquired over the mass range of 350–1800 and for product ion *m/z* 100–1800. Ions observed in the TOF-MS scan exceeding a threshold of 100 counts and a charge state of +2 to +5 were set to trigger the acquisition of product ion, *m/z* spectra of the resultant 20 most intense ions. The data were acquired and processed using Analyst TF1.5.1 software (AB Sciex). Spectra were deisotoped, and peaks with a local signal to noise ratio greater than 5 were picked and searched by local Mascot v. 2.1 (Matrix Science, Boston, MA) against a database of protein sequences derived from the *C. aspersum* protein database. The database was composed to represent a search criterion that identified conservation of active peptides.

Search parameters were as follows: trypsin digestion, fixative modification including carbamidomethylation, and variable modifications including amidation, methionine oxidation, conversion of glutamine to pyroglutamic acid, and deamidation of asparagine. Precursor mass error tolerance was set to 20 ppm, and fragment ion mass error tolerance was set to 0.1 Da. The maximum expectation value for accepting individual peptide ion scores [ $-10 \cdot \log(p)$ ] was set to  $\leq 0.01$ , where  $p$  is the probability that the observed match is a random event.

**In Vitro Assay 4: Synthetic Peptides**—A peptide corresponding to SEEDGFKYDDIDDVEAESENDERHVD of was selected based on MS analysis data of bioactive fractions. This peptide was synthesized by ChinaPeptides (Shenzhen, China), purified to 95% purity, and verified by RP-HPLC and LC-MS. The conditions under which the peptide was assayed followed those described in *in vitro* assays 1–3. Synthetic LDA was applied at 10  $\mu\text{M}$  peptide/dose (final concentration, 0.2  $\mu\text{M}$ ).

**LDA Comparative Amino Acid Sequence Identification and Phylogenetic Analysis**—To identify the LDA gene sequence, we searched the mollusk protein data set with BLASTX using the SEEDGFKYDDIDDVEAESENDERHVD peptide sequence as a query with an  $e$  value threshold of  $1 \times 10^{-6}$ . After identifying the putative *Ca*-LDA precursor, we then performed the reciprocal search with TBLASTN against the *T. pisana* brain transcriptome (22) using the *Ca*-LDA as a query with the same  $e$  value threshold. The expressed sequence tags database and BLASTp were queried to identify other putative LDA precursors from mollusks in the National Center for Biotechnology Information. One-directional best hits were declared for each query if only a single BLAST result was obtained or the ratio of the BLAST score of the second best hit to the BLAST score of the first best hit was less than 0.7. After obtaining putative LDA gene sequences from mollusks, SignalP 4.0 or Predisi was used to identify protein-coding genes containing a predicted signal peptide sequence (23). NeuroPred was employed to predict cleavage sites, post-translational modifications, and bioactive peptide products. Schematic diagrams of protein domain struc-

tures were prepared using Domain Graph (DOG, version 2.0) software. The protein three-dimensional model of LDA was built using the Assisted Model Building with Energy Refinement (AMBER) 14 program, in which the molecular dynamic simulations were sampled every picosecond for a total of 250 nanoseconds. A LDA phylogenetic tree was constructed on the MEGA 6.0 platform using the maximum likelihood test with the Jones-Taylor-Thornton method model with 1000 bootstrap replicates (24). Multiple sequence alignment schematics were generated with ClustalW and illustrated using LaTeX's TeX-Shade package (25).

**Gross Morphology of the Reproductive System, Histology, and Scanning Electron Microscopy (SEM) of Mucous Glands**—*C. aspersum* (body mass, 15–35 g) and *T. pisana* (body mass, 4–8 g) were anesthetized with an injection of 0.5 M  $\text{MgCl}_2$ . After 5 min, they were sacrificed by crushing the shell. Remnant shell was removed along with other nonessential organs, and the whole reproductive system was extracted carefully in 4% paraformaldehyde in 0.1 M PBS (pH 7.0). Overall morphology of the reproductive system was captured through individual micrographs using a Nikon Light microscope fitted with NIS Elements software package and merged using Photoshop (Adobe).

For light microscopy, dissected mucus glands were fixed in 4% paraformaldehyde overnight at 4 °C and subsequently dehydrated in ascending concentrations of ethanol for 30 min each. This was cleared with two changes of xylene and a final xylene-paraplast mixture for 30 min. Following paraffin wax embedding, 5- $\mu\text{m}$  serial sections were prepared with a rotary microtome and floated on super frosted slides and processed for hematoxylin and eosin staining as previously described (26). Micrographs were taken using a Nikon Eclipse Light microscope fitted with the NIS Elements software package and prepared for presentation using Photoshop (Adobe). Low power micrographs were stitched together based on analysis of overlapping parts of high power micrographs.

For SEM, isolated mucous glands from *C. aspersum* and *T. pisana* adults were washed with SMT solution (250 mM sucrose, 2 mM  $\text{MgCl}_2$ , and 10 mM Tris-HCl, pH 7.4), and fixed in a solution of 4% glutaraldehyde plus 2% paraformaldehyde in Millonig's buffer (pH 7.2) at 4 °C for 1 h, washed with three changes of the same buffer for 5 min each, to remove fixatives, followed by postfixation in 1% osmium tetroxide in the same buffer for 30 min, and washed again in three changes of the same buffer. After fixation, specimens were dehydrated in increasing concentrations of ethanol at 50, 70, 90, and 95% for 10 min each and three times at 100% for 15 min each. Specimens were dried in a Hitachi HCP-2 critical point drying machine, using liquid  $\text{CO}_2$  as a transitional medium. They were mounted on aluminum stubs with carbon paint adhesive and coated with gold in a Hitachi ion sputtering apparatus (E5000) for 1 min. The specimens were examined by a Hitachi S-2500 scanning electron microscope at 15 kV.

**Immunofluorescent Localization of LDA Precursor in Mucous Glands**—A polyclonal antibody was generated against the LDA precursor protein using keyhole limpet hemocyanin-coupled peptides to RLDKFGFSGGI-amide (Genscript). Mucous glands were isolated from *C. aspersum* and *T. pisana* and fixed

## Allohormone in Helicid Snails

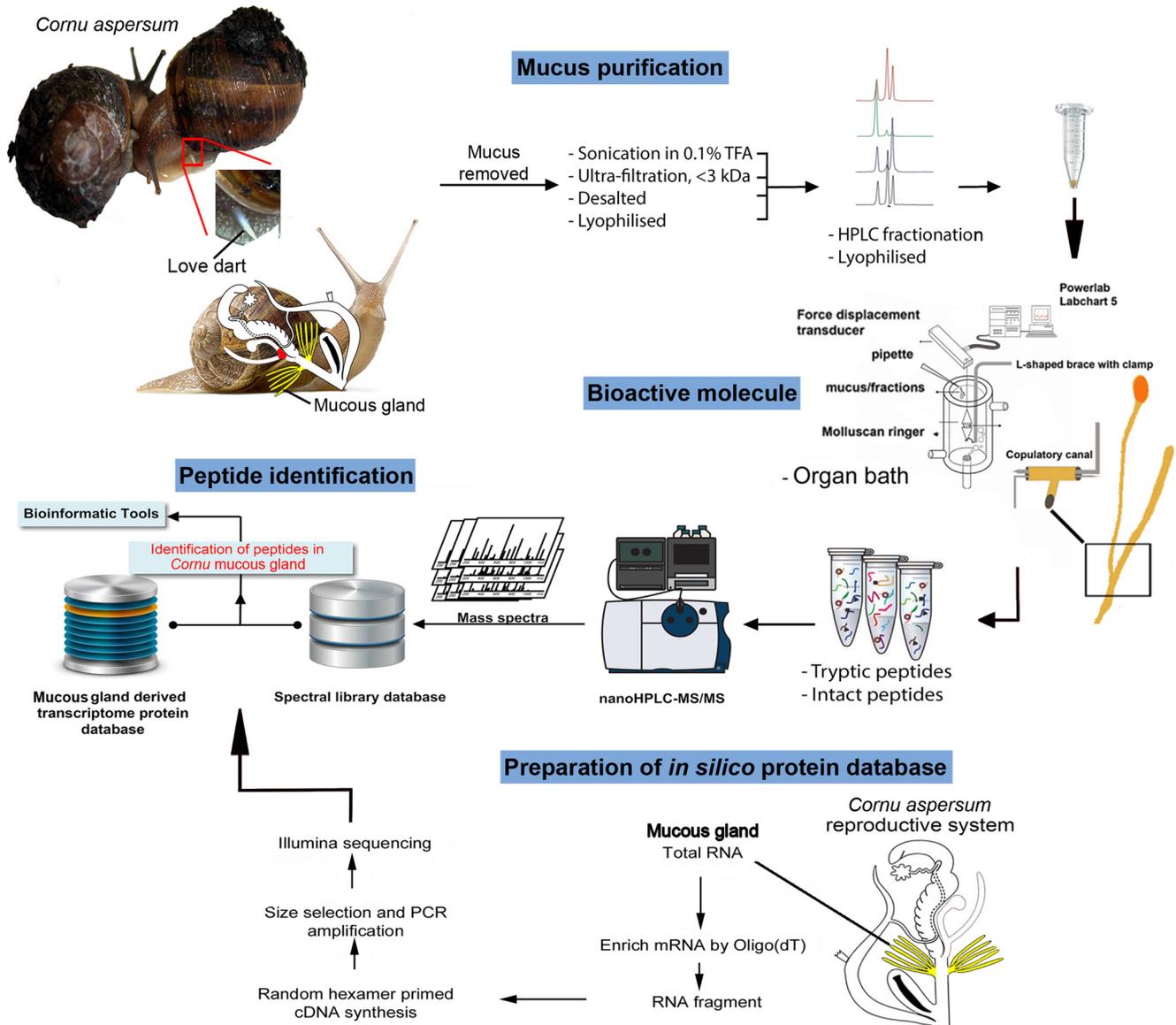


FIGURE 1. **Workflow for identification of *C. aspersum* love dart allohormone.** Shown are a copulating pair of *C. aspersum* showing a love dart and the reproductive system, highlighting the mucous glands (yellow) and site of allohormone activity (red).

in 4% paraformaldehyde overnight at 4 °C, before dehydration in ascending concentrations of ethanol for 30 min each, cleared in xylene three times, infiltrated, and embedded in paraffin. Serial transverse sections of the tissues were cut at 5- $\mu\text{m}$  thickness using a microtome. Sections were then deparaffinized in xylene and rehydrated in descending concentrations of ethanol. Subsequently, sections were incubated in 0.1% glycine in 0.1 M PBS for 30 min and washed three times with 0.1 M PBS containing 0.4% Triton X-100 (PBST). Nonspecific binding was blocked in 4% normal goat serum in PBST for 2 h, followed by incubation with the primary antibody (anti-peptide) at a dilution of 1:1000 in blocking solution at 4 °C overnight. Sections were then washed three times with PBST and incubated for 2 h with Alexa Fluor 488-conjugated goat anti-mouse IgG (Santa Cruz) at room temperature. After washing with PBST, the nuclei were stained with DAPI (Santa Cruz) for 10 min. Finally, sections were washed with PBST and mounted with VECTA shield fluorescent mounting medium

(Molecular Probes, Brisbane, Australia) before viewing under a confocal laser scanning microscope (Nikon) and prepared for presentation using Photoshop (Adobe). For negative controls, tissues were processed by the same protocol, but preimmune mouse serum was used instead of primary antibody.

### Results

*Mucous Glands Contain a Biomolecule That Stimulates Contraction of the Copulatory Canal*—Our study integrated multiple levels of analysis, including transcriptomic, proteomic, functional, and comparative analysis (Fig. 1). Digitiform mucous glands were collected from *C. aspersum* in which courtship, dart shooting, and copulation had been observed. Collected mucus was semipurified through size filtration into two fractions of protein molecular masses of  $>3$  and  $<3$  kDa and then tested on the T-junction of the copulatory canal, the region where the copulatory canal forks with the bursa tract

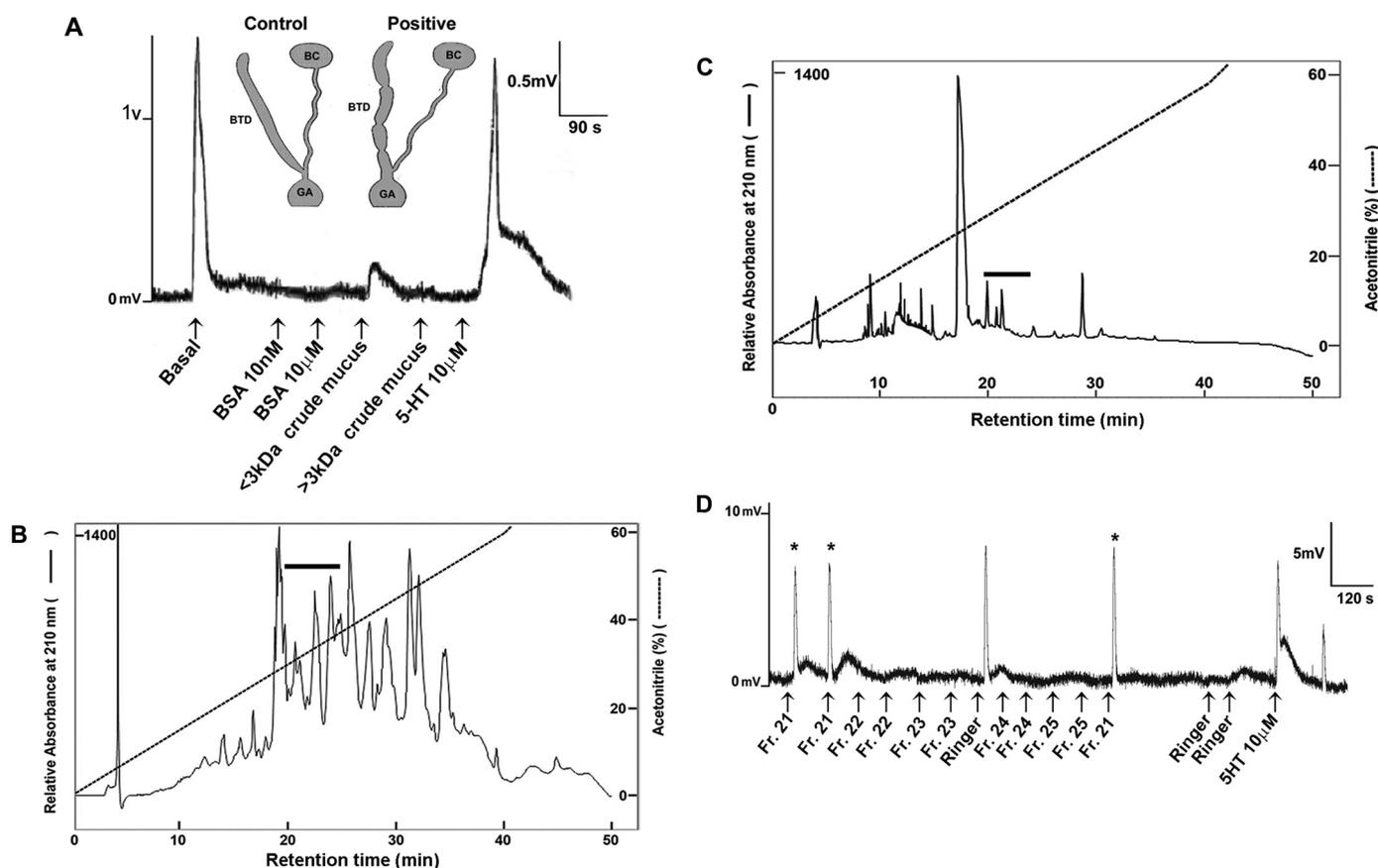


FIGURE 2. Identification of mucous gland-associated allohormone fraction in *C. aspersum*. *A*, representative *in vitro* assay showing the effect of mucus extracts on contraction of the copulatory canal, which forks into two distinctive functional regions: the bursa tract diverticulum (BTD) and bursa copulatrix (BC). Delivery of each extract was issued at approximately every 90 s at a dose of 20  $\mu$ l/extract. Extracts included negative controls of two levels of BSA, crude mucous gland extracts of <3 kDa and crude mucus extracts of >3 kDa, and the positive control serotonin (5-HT). Contractions (positive responses) are recorded as mV and compared with basal contractions and stimulated contractions induced by 5-HT. The schematics were adapted from Pomiankowski and Reguera (42). *B*, representative RP-HPLC chromatogram of a total *C. aspersum* mucous gland extract. The absorbance was monitored at 210 nm, and the gradient line (dotted line) shows the percentage of acetonitrile. The black bar shows the region that is bioactive in *in vitro* contractility assays (supplemental Table S1). *C*, RP-HPLC chromatogram of the *C. aspersum* mucous gland extract semipurified previously by a 2-ml Ultracel with a <3-kDa cutoff. The absorbance was monitored at 210 nm, and the gradient line (dotted line) shows the percentage of acetonitrile. The black bar shows the region that is bioactive in *in vitro* contractility assays (supplemental Table S1). *D*, representative *in vitro* assay showing the effect of RP-HPLC fractions 21–25. Delivery of each fraction was issued approximately every 120 s at a dose of 10  $\mu$ l/fraction. Contractions (positive responses) are recorded as mV and compared with basal contractions and those induced by 5-HT. Asterisk indicates positive response following fraction treatment.

diverticulum and bursa copulatrix (Fig. 2A). At the point in which basal contractility was observed, test preparations were administered. In response to <3-kDa mucus extract, the copulatory canal showed a sustained contraction over  $\sim$ 80 s (Fig. 2A and supplemental Table S1). Although amplitudes for <3-kDa extracts were not as pronounced as basal contractility, these were significant compared with negative controls (BSA) and >3-kDa mucus extract (Fisher's exact test:  $p = 0.00001$ ). The positive control (serotonin, 5-HT, 10  $\mu$ M) stimulated contractile activity consistent with that of the basal contraction, although 5-HT-induced contraction was generally longer in duration.

To elucidate the molecular identity of the <3-kDa mucus-associated secreted product(s) that did stimulate copulatory canal contractility, RP-HPLC was performed on total extracts prepared from *C. aspersum* mucous glands (Fig. 2B). Pooled 5-min fractions were lyophilized and reconstituted in water for systematic testing for contraction. Only the pooled HPLC fractions 21–25 consistently stimulated contractility ( $n = 3$ , 100%;  $p = 0.0047$ ). Subsequent *in vitro* analysis of the individual

fractions purified on the same gradient (Fig. 2C) showed that only fraction 21 (chi-square test:  $\chi^2 = 7.20$ ,  $p = 0.0119$ ;  $n = 8$ ) elicited contractility of the copulatory canal (Fig. 2D and supplemental Table S1).

To identify the mucous gland peptides, Illumina next generation transcriptome sequencing was first performed on *C. aspersum* mucous gland RNA (Fig. 1). This produced 58,717,616 raw reads, which after *de novo* sequence assembly into a mucous gland transcriptome resulted in 186,132 contigs and 57,750 unigenes. Prior to assembly, raw reads were trimmed of adapters and deposited into the GenBank<sup>TM</sup> sequence read archive database (SRX957716). After trimming, clean read length of unigenes was an average of 443 bp, and a total of 15,233 unigenes were found to have  $e$  values  $\leq e10$  (51.0%) when compared with the GenBank<sup>TM</sup> nonredundant database using BLASTx. An *in silico* mucous gland protein database was derived from this transcriptome, whereby individual transcripts were translated to ORFs and used for identification of genes encoding mucous gland peptides, including the LDA.

**TABLE 1**

Summary of proteins with BLASTp match (e value  $10^{-6}$ ) in mucus HPLC fraction 21

The amino acid sequences are shown in the supplemental File S1.

Accession numbers	No. of peptides	Description
Unigene32781_HaMG	5	Buccalin precursor
CL5959.Contig2_HaMG	5	Non-neuronal cytoplasmic intermediate filament protein
CL81.Contig1_HaMG	5	Actin
Unigene1286_HaMG	4	Protein disulfide-isomerase A3
Unigene25454_HaMG	2	Transgelin-2
CL6679.Contig3_HaMG	4	Calreticulin
CL321.Contig1_HaMG	4	78-kDa glucose-regulated protein precursor
Unigene31995_HaMG	2	Non-neuronal cytoplasmic intermediate filament protein
CL1651.Contig1_HaMG	3	Glucose-regulated protein 94
Unigene33640_HaMG	2	26S proteasome non-ATPase regulatory subunit 14
CL3127.Contig1_HaMG	2	Cholinesterase-like
Unigene33436_HaMG	1	Actin I
Unigene28810_HaMG	1	Putative polyadenylate-binding protein 1
CL378.Contig1_HaMG	1	Polyadenylate-binding protein 4
CL4509.Contig2_HaMG	2	Elongation factor-1 $\delta$ bisoform 1
CL4509.Contig1_HaMG	2	Elongation factor 1 $\beta$ -like
Unigene31980_HaMG	1	Nucleoside diphosphate kinase B
CL6817.Contig1_HaMG	2	Arginine kinase
Unigene34625_HaMG	1	Fructose-biphosphatealdolase
CL2802.Contig2_HaMG	1	Myosin regulatory light chain
Unigene32295_HaMG	1	Eukaryotic translation elongation factor 1 epsilon-1
CL5683.Contig2_HaMG	1	FMRF-amide neuropeptide
Unigene32511_HaMG	1	Predicted: myomodulin neuropeptides 1-like
Unigene33563_HaMG	1	Endoplasmic reticulum protein ERp29
Unigene26477_HaMG	1	Predicted: endoplasmic reticulum resident protein 44-like
CL5172.Contig2_HaMG	1	Heat shock protein cognate 5 partial
CL7404.Contig6_HaMG	1	HSP70
CL6230.Contig1_HaMG	1	Heat shock protein 70
CL6830.Contig1_HaMG	1	Heat shock cognate protein 70
Unigene2750_HaMG	1	Predicted: muscle M-line assembly protein unc-89-like, partial
Unigene20663_HaMG	1	Predicted: calcium-regulated heat stable protein 1-like
Unigene24727_HaMG	1	Putative ribosomal protein
Unigene32909_HaMG	2	Peptidylprolyl cis-trans isomerase B
CL3377.Contig1_HaMG	1	Heterogeneous nuclear ribonucleoprotein L
Unigene4188_HaMG	1	FFamide
Unigene8400_HaMG	1	Ribosomal protein rpl38
Unigene34590_HaMG	1	Predicted protein
Unigene26134_HaMG	1	Chain A placopecten striated muscle myosin
Unigene19980_HaMG	1	Predicted: 10-kDa heat shock protein, mitochondrial-like
Unigene34509_HaMG	1	Predicted: phosphoenolpyruvatephosphomutase-like
Unigene32934_HaMG	2	FVRIamide neuropeptide precursor
Unigene33965_HaMG	1	Ribosomal protein L10a
CL5866.Contig1_HaMG	1	Arginase type I-like protein
CL331.Contig3_HaMG	1	Acetylcholinesterase
CL331.Contig2_HaMG	1	Cholinesterase
CL331.Contig1_HaMG	1	Acetylcholinesterase putative
Unigene24585_HaMG	1	Myosin-VI
CL2369.Contig2_HaMG	1	Calumenin-B
CL1207.Contig1_HaMG	1	Predicted: glutaryl-CoA dehydrogenase, mitochondrial-like
CL3253.Contig1_HaMG	1	Paramyosin
CL5457.Contig2_HaMG	1	Matrix metalloproteinase-21
CL6000.Contig1_HaMG	1	Predicted: ATP synthase subunit $\beta$ , mitochondrial-like
Unigene35379_HaMG	1	Enolase-phosphatase E1
CL1213.Contig1_HaMG	1	DnaJ-like protein subfamily B member 11
CL6818.Contig1_HaMG	1	Protein disulfide-isomerase A6
Unigene32024_HaMG	1	Protein <i>Xenopus</i> (Silurana) <i>tropicalis</i>
CL6830.Contig1_HaMG	1	Heat shock cognate protein 70
Unigene33735_HaMG	1	Sphingomyelin phosphodiesterase 4

*Mucous Glands Contain Numerous Peptides, Including a Love Dart Allohormone*—To find the active fraction in the transcriptome, mucus HPLC fraction 21 (LDA) was repurified by RP-HPLC to resolve 14 subfractions. These were then analyzed by MALDI-TOF/TOF-MS, followed by matching with our mucous gland protein database. A total of 57 precursor proteins were identified that showed BLASTp matches (Table 1, supplemental Table S2, and supplemental File S1). These include matches for proteins previously defined as neuropeptides in other mollusks, including buccalin, FMR-Famide, myomodulin, FFamide, and FVRIamide. An additional 55 proteins did not match with any other protein in

the NCBI databases (supplemental Table S3). From HPLC fraction 21, the 14 resolved subfractions were individually tested for *in vitro* contractility. Fractions 7 and 12 were consistently bioactive (100% as triplicate,  $p < 0.01$  verified by chi-square test:  $\chi^2 = 9.60$ ; Fig. 3A and supplemental Table S1) over other fractions that showed as not meeting the threshold criterion to induce contractions, *i.e.* a contraction height near equal to the baseline contractile level. Following MALDI-TOF/TOF MS analysis of fractions, we identified the same the amino acid sequence SEEDGFKYDDI-DDVEAESENDRHVD (2856 Da), in both bioactive fractions 7 and 12, although it appeared to be present at varying concentrations throughout most subfractions (Fig. 3B). These varia-

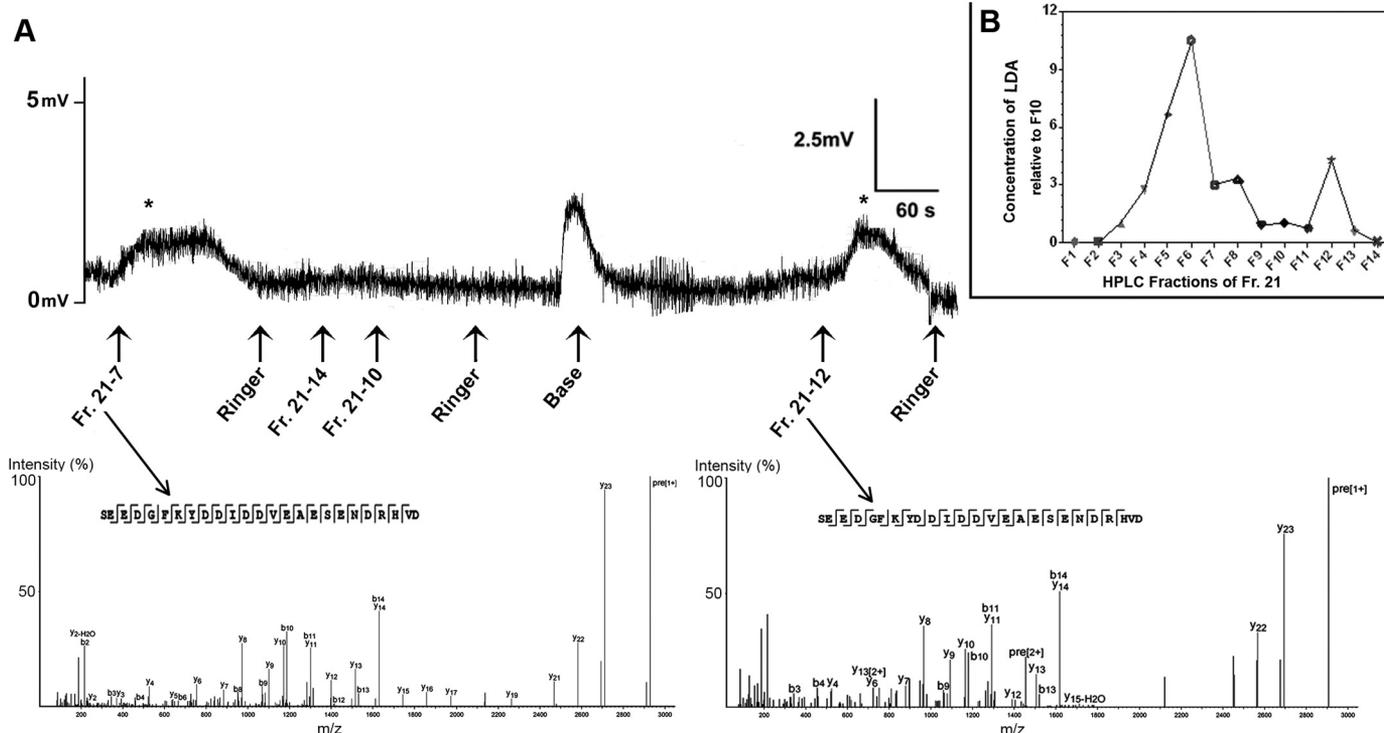


FIGURE 3. Identification of *C. aspersum* love dart allohormone. *A*, a representative *in vitro* organ bath assay shows effect of fraction 21 repurifications on bursa duct (see supplemental Table S1 for results). Asterisk indicates positive response following fraction treatment. Mass spectra obtained from different bioactive fractions showed the identical peptide sequence, named love dart allohormone (LDA). *B*, relative concentrations of LDA within fraction 21 repurifications, from fractions 1–14. *Fr.*, fraction.

tions thus likely account for the inconsistent nature of contraction of subfractions in organ bath assays.

**LDA Is Derived from the Mollusk Buccalin Precursor Protein in Helicid Snails**—The transcriptome data revealed that the identified *C. aspersum* LDA peptide is encoded by a precursor protein of 235 amino acids, which also includes a signal peptide and dibasic cleavage sites to release up to 6 amidated peptides and 3 nonamidated peptides (including LDA) (Fig. 4A). A representative tertiary structure for the LDA peptide is shown that occurred at 514.415 ns into the 700-ns molecular dynamic simulation (supplemental Fig. S1). Its structure primarily consists of an  $\alpha$ -helix, a turn, and a random coil.

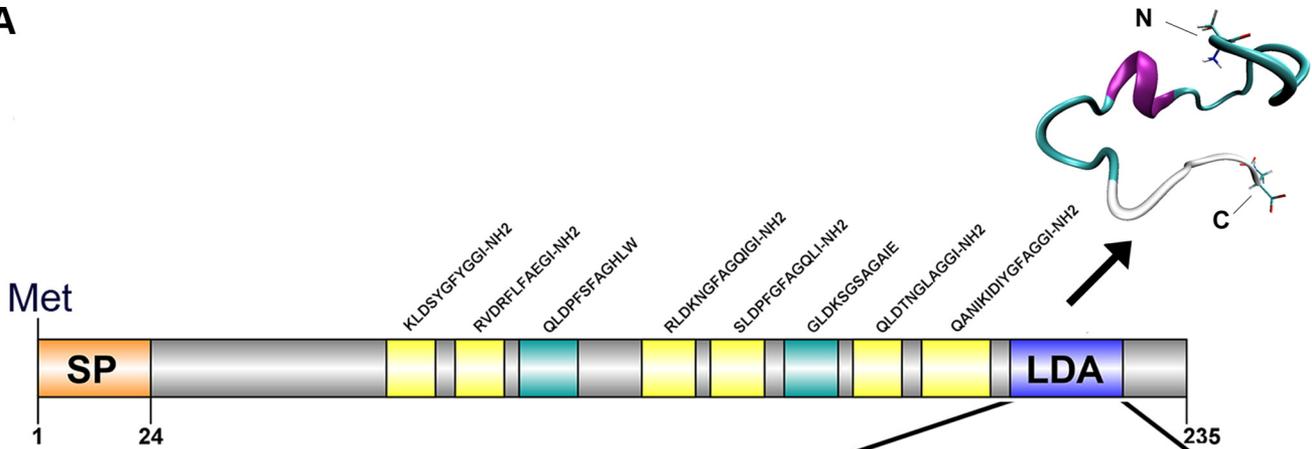
BLASTp analysis provided strong evidence that the *C. aspersum* LDA precursor is a homolog of buccalin precursors that have previously been identified in several mollusks, including the sea slug (*Aplysia californica*) and the freshwater pond snail (*Lymnaea stagnalis*) (27, 28) (Fig. 4B). A full-length precursor comparison indicates a similar precursor organization, particularly of 8–9 highly conserved amidated peptide repeats (supplemental Fig. S2). However, multiple amino sequence alignment of the *C. aspersum* LDA peptide with the corresponding region of representative aquatic and land mollusks indicates conservation within the helicid snails but little conservation with other mollusks (Fig. 4B). Phylogenetic analysis indicates that the gastropods and bivalves of have buccalin-like precursors, whereas the cephalopods do not (Fig. 4C).

To confirm that the identified peptide was responsible for the induced contractions, a synthetic *C. aspersum* LDA was tested in the *in vitro* contraction assay, and results are shown in Fig. 5A. Synthetic LDA (10  $\mu$ M) could elicit contraction on 95%

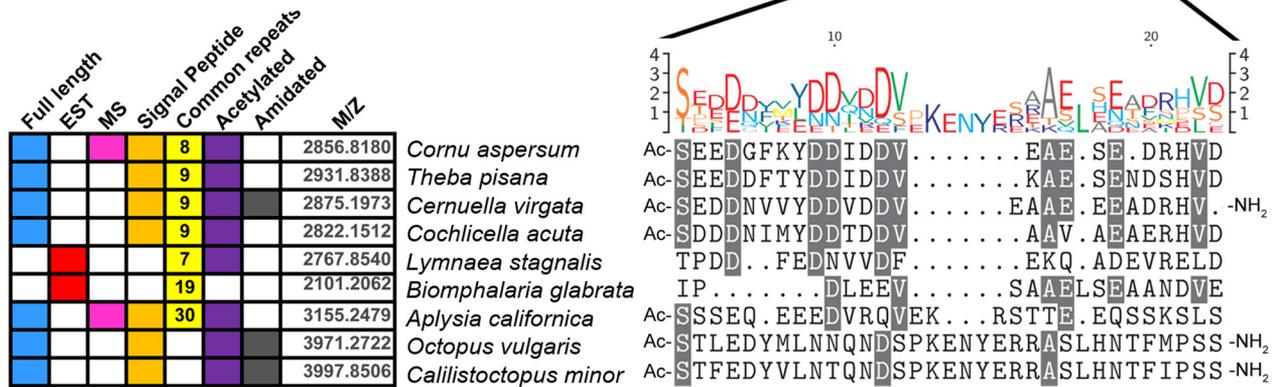
of times tested (chi-square test:  $\chi^2 = 4.09$ ,  $p < 0.05$ ) (Fig. 5B). This confirmed that LDA stimulates copulatory canal contraction, implicating this peptide as one of the molecules mediating paternity success in helicids that can now be tested experimentally.

**LDA Distributed throughout Mucous Glands of Helicid Snails**—LDA immunoreactivity was used to investigate the spatial expression of LDA within the mucous glands of *C. aspersum* and *T. pisana*, another helicid snail. Typical of helicids, their reproductive systems are composed of three major divisions (Fig. 6, A and B): 1) a hermaphroditic part, which includes the ovotestis and hermaphroditic duct; 2) a female part, which includes the oviduct proper, the fertilization pouch, spermatheca, albumen gland, egg membrane gland, the egg mass membrane gland, and the vagina, with which the duct of the bursa copulatrix is closely associated; and 3) a male part, which includes the penis and associated organs, penis retractor muscle, dart sac, and mucous glands. The male and female parts of the reproductive system of both species are quite distinct morphologically, yet they are intricately entwined, except for the more distal half of the copulatory organ and bursa copulatrix. The mucous glands in *C. aspersum* are complexly branched with a finger-like appearance, as highlighted by SEM that also shows a uniform smooth surface with no distinct external features (Fig. 6C). However, in *T. pisana*, mucous glands under SEM appear acinous with distinct irregular lobular structures consisting of a number of convoluted ridges (hillocks) separated by serrated grooves (Fig. 6D). Mucous glands of both species are fixed at the junction that intersects the dart sac.

A



B



C

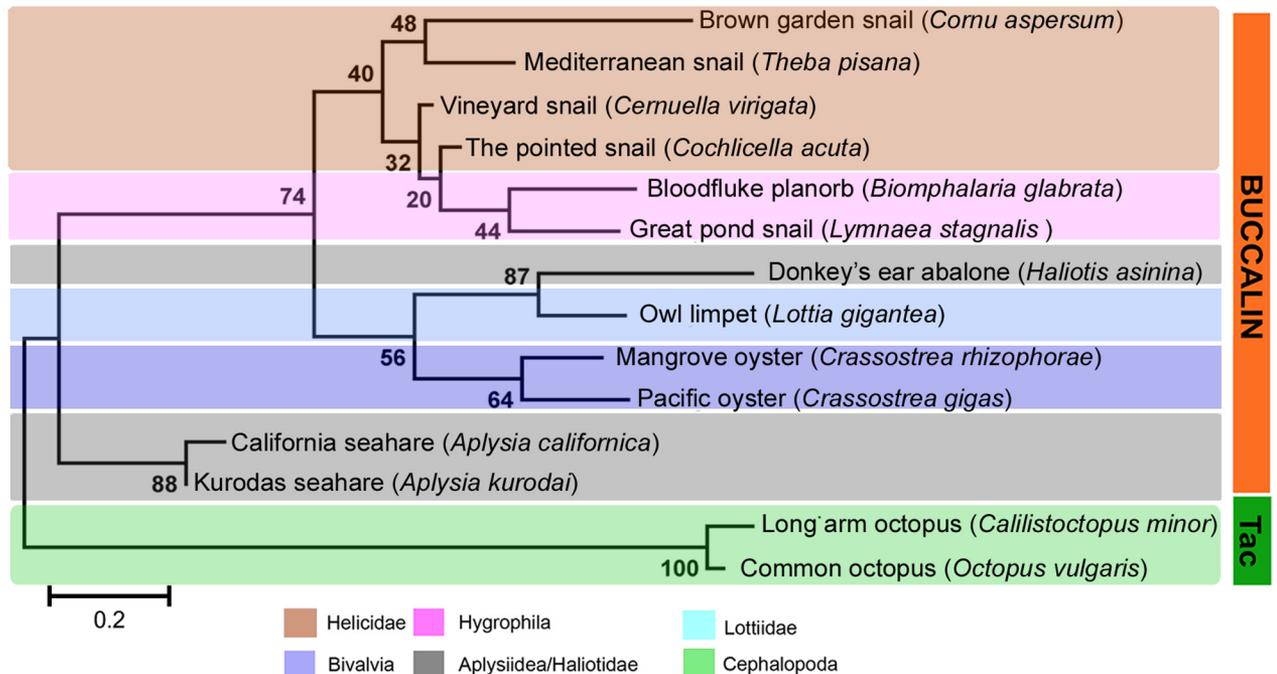


FIGURE 4. LDA characterization in *C. aspersum*, comparative analysis and synthetic bioassay. A, full-length precursor schematic of the LDA precursor. SP, signal peptide; yellow, amidated peptides; green, nonamidated predicted peptides; blue, LDA. Vertical lines represent dibasic cleavage sites, unless shown as KRR. A protein model of the LDA peptide assembled by molecular dynamics simulation shows  $\alpha$ -helix, Ile<sup>11</sup>-Glu<sup>15</sup>; Turn, Ser<sup>1</sup>-Glu<sup>10</sup> and Ala<sup>16</sup>-Asn<sup>20</sup>; and random coil, Asp<sup>21</sup>-Asp<sup>25</sup>. B, summary of the putative LDA-like precursors encoding putative full-length or partial-length precursors from genomic and expressed sequence tag (EST) libraries. For each LDA, the table also shows whether a full-length, expressed sequence tag ORF sequence, or MS evidence are available as well as whether common repeats (including numbers of peptides) occur if the peptides are amidated and acetylated, if a leader signal peptide is encoded, and the predicted molecular mass (*m/z*) of the LDA peptide. A comparative sequence alignment showing high amino acid identity between mollusk species that retain a putative LDA peptide is illustrated. The height of each letter of the logo is proportional to the observed frequency of the corresponding amino acid in the alignment column. C, phylogenetic analysis of the LDA-like precursors between species built using the maximum likelihood method (bootstrap *n* = 1000). Most appear to be homologs of buccalin, but two cephalopod peptides that are tachykinin-like (*Tac*) are also included. For accession numbers and protein sequences, see supplemental File S2.

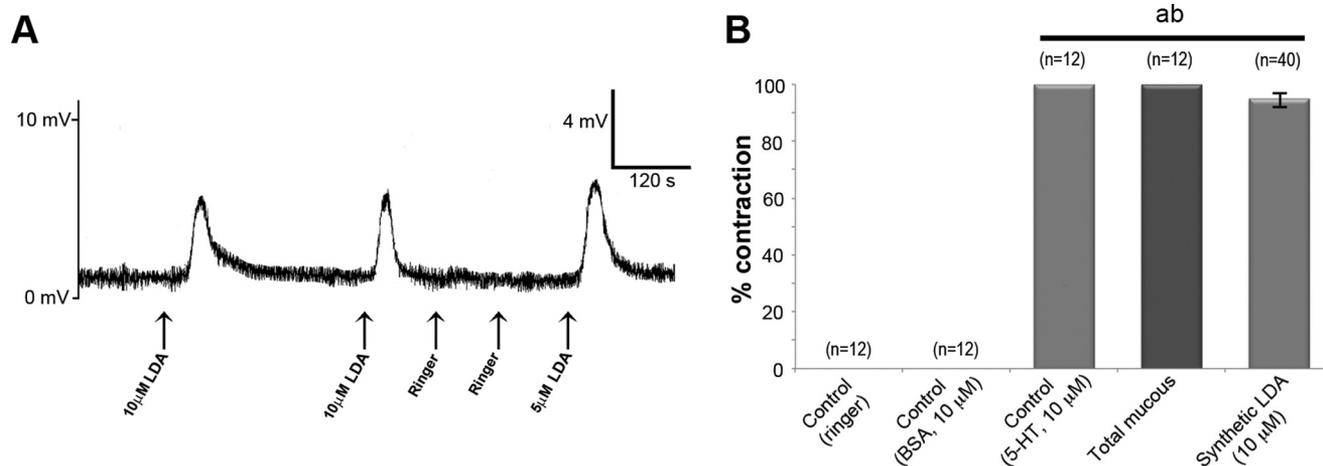


FIGURE 5. **A** representative *in vitro* organ bath assay shows the effects of peptides on the copulatory canal. **A**, contractions (positive responses) following application of 10  $\mu$ M LDA are recorded as mV. **B**, graph provides a summary of the *in vitro* organ bath assay with *n* values shown over each bar. <sup>ab</sup>, significantly different from controls ( $p < 0.05$ ); <sup>a</sup>, Ringer; <sup>b</sup>, BSA.

In *C. aspersum*, the histology of the mucous gland revealed multiple eosinophilic goblet-like cells located on the outer periphery, with a central core full of densely packaged basophilic staining mucus (Fig. 6E). This is in addition to two types of mucus-secreting cells (msc); msc-I, mucus-deficient or empty cells, and msc-II, mucus-bearing cells (Fig. 6E, top right panel). Notably at the tip of each mucous gland are stores of basophilic mucus surrounded by supporting cells that possess loosely dispersed nuclei (Fig. 6E, bottom right panel). The *T. pisana* mucous gland contains aggregates of enclosed glandular pockets supported by goblet-like cells containing mucus (Fig. 6F, inset panel, left top). At the tip of the mucous gland is a dense mass of eosinophilic mucus (Fig. 6F, inset panel, left bottom).

An antibody raised against the LDA precursor was used to localize expression within the mucous glands (Fig. 6, G and H). Abundant immunoreactivity was present in both species, particularly within interstitial spaces, the mucous gland duct, and the outer and inner trabeculae that are comprised of striated muscle, connective tissue, and supporting cells. However, no LDA immunoreactivity was observed within mucus-secreting cells themselves.

## Discussion

The snail's love dart activity has been documented in the literature as far back as the mid-17th century, and love dart-possessing snails were known to the ancient Greeks, probably influencing the creation of the cupid myth. However, the exact molecular identity of the active substances that the dart transfers has remained a mystery for many years. Here, we have identified the peptide that is produced in the helicid mucous glands and that is responsible for the previously reported conformational changes in the reproductive tract (13).

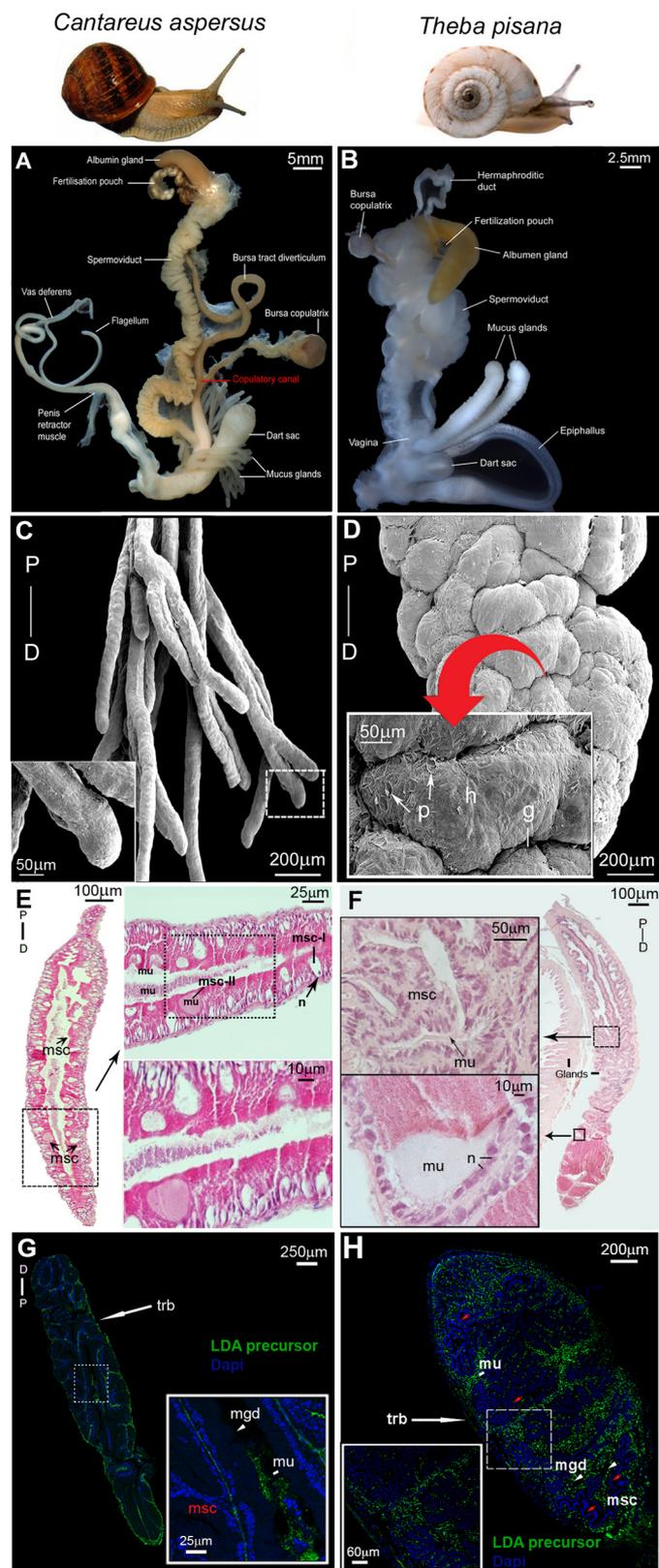
We used the garden snail *C. aspersum* to elucidate the components of the mucous gland and more specifically the molecular identity of the LDA. *C. aspersum* was most appropriate for this study because it had represented the helicid model for defining whether the dart achieves its effect by a mechanical action or by transferring a bioactive substance; injections of

*C. aspersum* mucus from the mucous gland more than doubled paternity relative to injections of saline (10). The use of *C. aspersum* did also provide us with a relatively large bursa duct that could readily facilitate the experimental hook-up of the T-junction of the copulatory canal (forks with the bursa tract diverticulum and bursa copulatrix) for contraction bioassay. The physiological basis for LDA bioactivity is through muscular stimulation, resulting in inhibition of sperm transfer to the bursa copulatrix (13). To confirm that the identified peptide was responsible for the induced contractions, a synthetic *C. aspersum* LDA was tested in the *in vitro* contraction assay, confirming that LDA stimulates copulatory canal contraction, implicating this peptide as one of the molecules mediating paternity success in helicids that can now be tested experimentally.

The next generation sequencing transcriptome derived from the *C. aspersum* mucous gland enabled the rapid identification of purified bioactive peptide. This transcriptome, containing 186,132 contigs and 57,750 unigenes was also helpful to determine the existence of numerous other peptides, in addition to the LDA. For example, we identified other peptides previously defined as neuropeptides in other mollusks, including FMRFamide, myomodulin, FFamide, and FVRLamide (29). These neuropeptides have been reported sporadically throughout molluscan literature, although primarily within the aquatic snails. In land snails, the FMRFa precursor protein has been identified through *in silico* search of the *T. pisana* CNS transcriptome data set (22) and is known to have a cardioexcitatory effect in other mollusks (30, 31). The myomodulin gene was also identified from the *T. pisana* CNS (22), and previously the peptide was identified in *C. aspersum* where 3 myomodulin-related peptides are distributed throughout 26 different snail tissues (highest levels occurring in certain male reproductive organs) (32). A synthetic myomodulin (pQLSMLRLamide) could modify the spontaneous rhythmic activity or the resting tone of several isolated muscular organs (32), although the T-junction of the copulatory canal was not tested.

Bioassay of mucus RP-HPLC separated into subfractions found that only two fractions were consistently bioactive for

## Allohormone in Helicid Snails



**FIGURE 6. Localization of LDA in mucous glands of *C. aspersus* and *T. pisana*.** A and B, *C. aspersus* (A) and *T. pisana* (B) reproductive systems, including the bursa copulatrix involved in degrading sperm, the copulatory canal (observed in *C. aspersus*), which is the major site that directs the closure of the aperture that allows for sperm storage and/or sperm degradation at the BC, and paired mucous glands with associated dart sac. C and D, scanning electron micrographs of one branch of the mucous gland pair of *C. aspersus* (C) and *T. pisana* (D). P, proximal; D, distal. Higher power micrographs are

bursa tract contraction, yet MS analysis had identified that the same peptide LDA sequence within all subfractions. One explanation may be that it was present at varying concentrations throughout most subfractions, thus likely accounting for the inconsistent nature of contraction in organ bath assays.

Our molecular analysis demonstrated that the helicid LDA is released from a buccalin-like precursor, a peptide-rich precursor that is also found in other mollusks. BLASTp analysis provided strong evidence that the *C. aspersum* LDA precursor is a homolog of buccalin precursors that have previously been identified in several mollusks, including the sea slug (*Aplysia californica*), the freshwater pond snail (*Lymnaea stagnalis*), and the land snail *T. pisana* (27, 28, 33). It is striking that two of the helicids clearly differ from the other two land snails in this comparison (*Ceruella virgata* and *Cochlicella acuta*). *Ceruella* has a relatively small dart and is considered to be more basal (in the family Hygromidae), as defined by the presence of an effuse shell aperture. By comparison, the pointed snail *Cochlicella* is also basal and has only rudimentary dart glands and no dart (17). Other peptides within the precursor protein include the buccalin peptides, which have been implicated in depressing muscle contractions in *Aplysia* and as neuromodulators given their widespread distribution in nervous tissue (34–36). This function is supported by the presence of the C-terminal GLamide, which is a feature also found in insect allotostatin A (37) and in the class of DLamides identified in the annelid *Platynereis demurilii* (38). The buccalin gene of *T. pisana* has widespread expression outside the mucous gland, including the CNS of aestivated (dormant in warm conditions) snails, where the buccalin-2 isoform is distributed within regions that control several physiological roles (33).

Other LDA precursor peptides are likely coinjected into the recipient but were not tested in this study. The dart mucus can also affect the mating behavior of the recipient snails. *C. aspersum* that have had their digitiform mucous glands surgically removed required more courtship time to reach copulation than do negative controls (39). This demonstrates that it is the mucus from the digitiform mucous gland, not the mechanical action of the dart, which affects courtship duration. Also, injections of mucous gland homogenate decreases courtship duration of sexually receptive snails, whereas gland homogenate increases the size of the recipient's genital eversion, and retards locomotion (39). In addition, the mucus appears to affect the sperm recipient in other ways, as demonstrated in the snail *Euhadra quaesita*, in which the mucus suppresses subsequent matings and promotes oviposition in their partners (40), suggesting that the LDA and/or other as yet uncharacterized peptides have other physiological functions. Similarly, seminal

shown (inset). In *T. pisana*, convoluted ridges/hillocks (h) are separated by serrated grooves (g) and pits (p) of ~50µm. E, hematoxylin and eosin histology of the *C. aspersum* mucous gland showing msc. Higher magnifications are shown as horizontal insets that demonstrate mucus-deficient cells (msc-I) with a distinct nucleus (n), and mucus-bearing cells (msc-II) rich with mucus (mu). F, hematoxylin and eosin histology of the *T. pisana* mucous gland showing a similar anatomical arrangement of msc with stores of mucus. Higher power micrographs are shown as insets. G and H, immunolocalization of LDA precursor in *C. aspersum* and *T. pisana* mucous glands. Both show immunoreactivity at the region of the outer trabeculae (trb) and in mucous gland ducts (mgd).

fluid peptides of *Drosophila melanogaster* include a “sex peptide” that appears to be a global regulator of reproductive processes in mated recipients, including altered fecundity and sexual receptivity, and can lead to a conflict of interest over the control of reproduction (41). We expect that further studies could now be established that focus on the physiological and behavioral effects of other LDA precursor peptides, as well as other peptides that we have identified in the mucus (see supplemental Tables S2 and S3).

In summary, we present further evidence for the importance of allohormone type transfer between organisms through the identification of the LDA in a helicid snail. Given the neuro-modulatory action of LDA, this opens up new opportunities for understanding the evolution of accessory gland products and alternative reproductive strategies.

**Author Contributions**—M. J. S. assisted with the experimental design, performed the experiments, analyzed results, and drafted the manuscript. T. W. carried out MS analysis and compiled the protein model. J. M. K. and K. B. S. conceived the idea and drafted the manuscript. S. F. C. conceived the idea, constructed the experimental design, performed experiments, analyzed results, and drafted the manuscript.

**Acknowledgments**—We thank Graham Hayes (Yorke Peninsula; *T. pisana*), and Cliff Wilson and Mary Page (Glasshouse Gourmet Snails; *C. aspersum*) for supplying the snails used in this study, and Dr. Praphaporn Stewart for providing SEM of the mucous glands. We gratefully thank Dr. Alun Jones (Institute for Molecular Bioscience, University of Queensland) for advice and assistance with the LC-MS/MS. We thank the (Australian) National Computational Merit Allocation Scheme for supercomputing facilities.

**Note Added in Proof**—In the original version of the manuscript that was published as a JBC Paper in Press on January 27, 2016, the high magnification insets of panels E–H in Fig. 6 did not correspond to the boxed regions featured in the lower magnification images. These errors have been corrected, and the corrections do not affect the interpretation of the results or the conclusions of the article.

## References

- Wyatt, T. D. (2009) Fifty years of pheromones. *Nature* **457**, 262–263
- Møller, A. P. (1998) Sperm competition and sexual selection. In *Sperm Competition and Sexual Selection* (Birkhead, T. R., and Møller, A. P., eds) pp. 55–90, Academic Press, San Diego, CA
- Birkhead, T. R., Hosken, D. J., and Pitnick, S. S. (2008) *Sperm Biology: An Evolutionary Perspective*, pp. 207–245, Academic Press, Sheffield, UK
- Charnov, E. L. (1979) Simultaneous hermaphroditism and sexual selection. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 2480–2484
- Scharer, L., and Pen, I. (2013) Sex allocation and investment into pre- and post-copulatory traits in simultaneous hermaphrodites: the role of polyandry and local sperm competition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120052
- Nakadera, Y., and Koene, J. M. (2013) Reproductive strategies in hermaphroditic gastropods: conceptual and empirical approaches 1. *Can. J. Zool.* **91**, 367–381
- Leonard, J. L. (2006) Sexual selection: lessons from hermaphrodite mating systems. *Integr. Comp. Biol.* **46**, 349–367
- Baur, B. (1998) Sperm competition in molluscs. In *Sperm Competition and Sexual Selection* (Birkhead, T. R., and Møller, A. P., eds) pp. 255–306, Academic Press, San Diego, CA
- Chung, D. J. D. (1987) Courtship and dart shooting behavior of the land snail *Helix aspersa*. *Veliger* **30**, 24–39
- Chase, R., and Blanchard, K. C. (2006) The snail's love-dart delivers mucus to increase paternity. *Proc. R. Soc. Lond. B Biol. Sci.* **273**, 1471–1475
- Adamo, S. A., and Chase, R. (1988) Courtship and copulation in the terrestrial snail *Helix aspersa*. *Can. J. Zool.* **66**, 1446–1453
- Hunt, S. (1979) The structure and composition of the love dart (gypsobelum) in *Helix pomatia*. *Tissue Cell* **11**, 51–61
- Koene, J. M., and Chase, R. (1998) Changes in the reproductive system of the snail *Helix aspersa* caused by mucus from the love dart. *J. Exp. Biol.* **201**, 2313–2319
- Leonard, J. L. (1992) The “love-dart” in Helicid snails: a gift of calcium or a firm commitment? *J. Theor. Biol.* **159**, 513–521
- Adamo, S. A., and Chase, R. (1990) Dissociation of sexual arousal and sexual proclivity in the garden snail, *Helix aspersa*. *Behav. Neural Biol.* **54**, 115–130
- Rogers, D., and Chase, R. (2001) Dart receipt promotes sperm storage in the garden snail *Helix aspersa*. *Behav. Ecol. Sociobiol.* **50**, 122–127
- Koene, J. M., and Schulenburg, H. (2005) Shooting darts: co-evolution and counter-adaptation in hermaphroditic snails. *BMC Evol. Biol.* **5**, 25
- Koene, J. M., and Chase, R. (1998) The love dart of *Helix aspersa* Muller is not a gift of calcium. *J. Moll. Stud.* **64**, 75–80
- Lind, H. (1973) The functional significance of the spermatophore and the fate of spermatozoa in the genital tract of *Helix pomatia* (Gastropoda, Stylommatophora). *J. Zool.* **169**, 39–64
- Koene, J. M., and ter Maat, A. (2001) “Allohormones”: a class of bioactive substances favoured by sexual selection. *J. Comp. Physiol. A* **187**, 323–326
- Bernard, A., and Bonnet, V. (1930) The mineral composition of haemolymph and study of physiological solution for the snail. *R. Soc. Biol. Paris* **103**, 1110–1120
- Adamson, K. J., Wang, T., Zhao, M., Bell, F., Kuballa, A. V., Storey, K. B., and Cummins, S. F. (2015) Molecular insights into land snail neuropeptides through transcriptome and comparative gene analysis. *BMC Genomics* **16**, 308
- Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282
- Beitz, E. (2000) TEXshade: shading and labeling of multiple sequence alignments using LATEX2 epsilon. *Bioinformatics* **16**, 135–139
- Kuanpradit, C., Stewart, M. J., York, P. S., Degnan, B. M., Sobhon, P., Hanna, P. J., Chavadej, J., and Cummins, S. F. (2012) Characterization of mucus-associated proteins from abalone (*Haliotis*): candidates for chemical signaling. *FEBS J.* **279**, 437–450
- Cropper, E. C., Miller, M. W., Tenenbaum, R., Kolks, M. A., Kupfermann, I., and Weiss, K. R. (1988) Structure and action of buccalin: a modulatory neuropeptide localized to an identified small cardioactive peptide-containing cholinergic motor neuron of *Aplysia californica*. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6177–6181
- Santama, N., Wheeler, C. H., Burke, J. F., and Benjamin, P. R. (1994) Neuropeptides myomodulin, small cardioactive peptide, and buccalin in the central nervous system of *Lymnaea stagnalis*: purification, immunoreactivity, and artifacts. *J. Comp. Neurol.* **342**, 335–351
- Morishita, F., Furukawa, Y., Matsushima, O., and Minakata, H. (2010) Regulatory actions of neuropeptides and peptide hormones on the reproduction of molluscs. *Can. J. Zool.* **88**, 825–845
- Higgins, W. J., Price, D. A., and Greenberg, M. J. (1978) FMRFamide increases the adenylate cyclase activity and cyclic AMP level of molluscan heart. *Eur. J. Pharmacol.* **48**, 425–430
- Price, D. A., and Greenberg, M. J. (1989) The hunting of the FaRPs: the distribution of FMRFamide-related peptides. *Biol. Bull.* **177**, 198–205
- Greenberg, M. J., Doble, K. E., Lesser, W., Lee, T. D., Pennell, N. A., Morgan, C. G., and Price, D. A. (1997) Characterization of myomodulin-related peptides from the pulmonate snail *Helix aspersa*. *Peptides* **18**, 1099–1106
- Adamson, K. J., Wang, T., Rotgans, B., Kruangkum, T., Kuballa, A. V., Storey, K. B., and Cummins, S. F. (2015) Genes and associated peptides

## Allohormone in Helicid Snails

- involved with aestivation in a land snail. *Gen. Comp. Endocrinol.* **pii**, S0016–6480(15)30013–7
34. Cropper, E. C., Miller, M. W., Vilim, F. S., Tenenbaum, R., Kupfermann, I., and Weiss, K. R. (1990) Buccalin is present in the cholinergic motor neuron B16 of *Aplysia* and it depresses accessory radula closer muscle contractions evoked by stimulation of B16. *Brain Res.* **512**, 175–179
  35. Vilim, F. S., Cropper, E. C., Rosen, S. C., Tenenbaum, R., Kupfermann, I., and Weiss, K. R. (1994) Structure, localization, and action of buccalin B: a bioactive peptide from *Aplysia*. *Peptides* **15**, 959–969
  36. Miller, M. W., Alevizos, A., Cropper, E. C., Kupfermann, I., and Weiss, K. R. (1992) Distribution of buccalin-like immunoreactivity in the central nervous system and peripheral tissues of *Aplysia californica*. *J. Comp. Neurol.* **320**, 182–195
  37. Veenstra, J. A. (2010) Neurohormones and neuropeptides encoded by the genome of *Lottia gigantea*, with reference to other mollusks and insects. *Gen. Comp. Endocrinol.* **167**, 86–103
  38. Conzelmann, M., and Jékely, G. (2012) Antibodies against conserved amidated neuropeptide epitopes enrich the comparative neurobiology toolbox. *Evodevo* **3**, 23
  39. Adamo, S. A., and Chase, R. (1990) The “love dart” of the snail *Helix aspersa* injects a pheromone that decreases courtship duration. *J. Exp. Zool.* **255**, 80–87
  40. Kimura, K., Shibuya, K., and Chiba, S. (2013) The mucus of a land snail love-dart suppresses subsequent matings in darted individuals. *Anim. Behav.* **85**, 631–635
  41. Gioti, A., Wigby, S., Wertheim, B., Schuster, E., Martinez, P., Pennington, C. J., Partridge, L., and Chapman, T. (2012) Sex peptide of *Drosophila melanogaster* males is a global regulator of reproductive processes in females. *Proc. Biol. Sci.* **279**, 4423–4432
  42. Pomiankowski, A., and Reguera, P. (2001) The point of love. *Trends Ecol. Evol.* **16**, 533–534

# SUPPLEMENTAL FILES

**for**

A 'love' dart all hormone identified in the mucous glands of hermaphroditic land snails

Michael J. Stewart, Tianfang Wang, Joris M. Koene, Kenneth B. Storey and Scott F. Cummins

## Supplementary files

**Figure S1.** Molecular dynamics simulation of LDA peptide.

**Figure S2.** Multiple amino acid sequence alignment of buccalin-like precursor proteins.

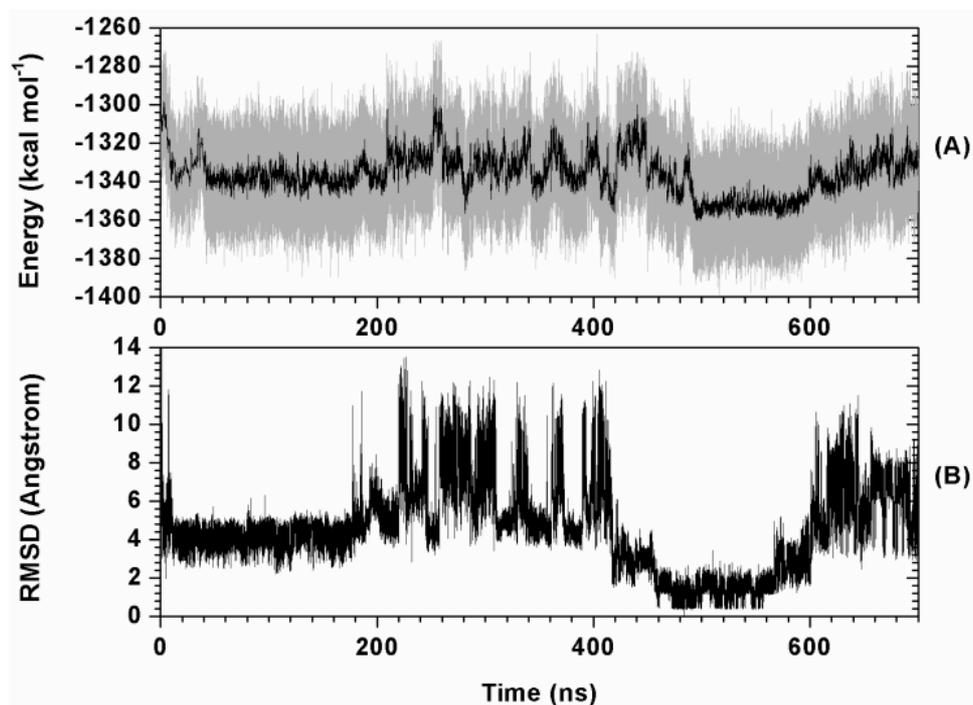
**Table S1.** Bioassays on the *C. aspersum* copulatory canal using organ bath.

**Table S2.** Full Summary of proteins with BLAST match (E-value 10<sup>-6</sup>) in RP-HPLC fraction 21.

**Table S3.** Summary of proteins with no BLASTp match in HPLC fraction 21.

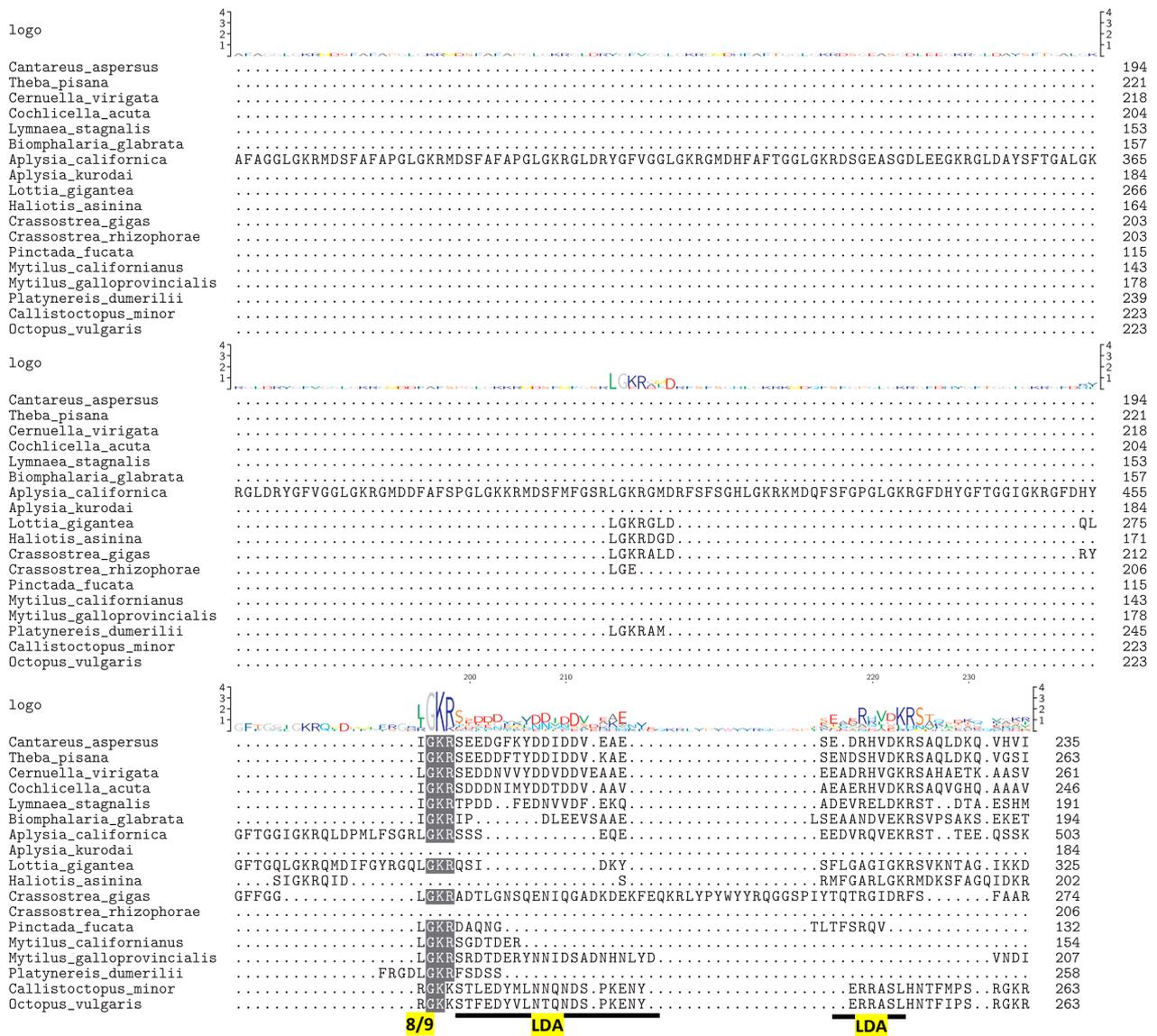
**File S1.** List of *Cornu aspersum* mucous gland precursor proteins derived from RP-HPLC fraction 21.

**File S2** – List of LDA precursor and buccalin-like precursors.



**Figure S1. Molecular dynamics simulation of LDA peptide.** (A) Potential energy of *SEEDGFKYDDIDDVEAESENDERHVD-OH* as a function of time during Molecular dynamic simulation (MD). The solid line is a running average over 100 ps. (B) Backbone root mean square deviation (RMSD) during the same MD, compared to the lowest-energy conformation.





**Figure S2. Multiple amino acid sequence alignment of buccalin-like precursor proteins.** Underscores and numbers reflect conserved repeats. For accession numbers see **File S1**.

**Table S1. Bioassays on the *C. aspersum* copulatory canal using organ bath.**

Assay	Stimulus	<i>n</i>	negative	positive
<b>1</b>	BSA 10 nM <sup>a</sup>	5	5	0
	BSA 10 μM <sup>b</sup>	5	5	0
	<3kDa crude mucus <sup>c</sup>	8	0	8 <sup>*abd</sup>
	>3kDa crude mucus <sup>d</sup>	8	8	0
	5-HT 10 μM <sup>e</sup>	8	0	8 <sup>*abd</sup>
	* <i>p</i> value<0.05 (0.00001)			
<b>2a</b>	BSA10μm <sup>a</sup>	3	3	0
	Ringer <sup>b</sup>	3	3	0
	1-5 <sup>c</sup>	3	3	0
	6-10 <sup>d</sup>	3	3	0
	11-15 <sup>e</sup>	3	3	0
	16-20 <sup>f</sup>	3	3	0
	21-25 <sup>g</sup>	3	0	3 <sup>*a-f,h,j</sup>
	26-30 <sup>h</sup>	3	3	0
	31-35 <sup>i</sup>	3	3	0
	36-40 <sup>j</sup>	3	3	0
	5-HT 10 μM <sup>k</sup>	3	0	3 <sup>*a-f,h,j</sup>
* <i>p</i> value<0.05 (0.004701)				
<b>2b</b>	Fr21 <sup>a</sup>	8	0	8 <sup>*bcdeg</sup>
	Fr22 <sup>b</sup>	3	3	0
	Fr23 <sup>c</sup>	3	3	0
	Fr24 <sup>d</sup>	3	3	0
	Fr25 <sup>e</sup>	3	3	0
	5-HT 10 μM <sup>f</sup>	3	0	3 <sup>*bcdeg</sup>
	Ringer <sup>g</sup>	3	3	0
	* <i>p</i> value<0.05 (0.011905)			
<b>3</b>	Ringer <sup>a</sup>	3	3	0
	BSA 10 μM <sup>b</sup>	3	3	0
	Sub-fraction 1 <sup>c</sup>	3	3	0
	Sub-fraction 2 <sup>d</sup>	3	3	0
	Sub-fraction 3 <sup>e</sup>	3	3	0
	Sub-fraction 4 <sup>e</sup>	3	3	0
	Sub-fraction 5 <sup>f</sup>	3	3	0
	Sub-fraction 6 <sup>g</sup>	3	3	0
	Sub-fraction 7 <sup>h</sup>	3	0	3 <sup>*a-g,i-l,n-o</sup>
	Sub-fraction 8 <sup>i</sup>	3	3	0
	Sub-fraction 9 <sup>j</sup>	3	3	0
	Sub-fraction 10 <sup>k</sup>	3	3	0
	Sub-fraction 11 <sup>l</sup>	3	3	0
	Sub-fraction 12 <sup>m</sup>	3	0	3 <sup>*a-g,i-l,n-o</sup>
	Sub-fraction 13 <sup>n</sup>	3	3	0
Sub-fraction 14 <sup>o</sup>	3	3	0	
5-HT 10 μM <sup>p</sup>	3	0	3 <sup>*a-g,i-l,n-o</sup>	
* <i>p</i> value<0.001 (0.00000014)				

\*significantly different based on Chi square tests

**Table S2. Full Summary of proteins with BLAST match (E-value 10<sup>-6</sup>) in HPLC Fraction 21.** Matches to known molluscan neuropeptides are highlighted in yellow. PTM = post translational modification. Amino acid sequences are shown in **File S1**.

Accession	-10lgP	#Peptides	PTM	Description
Unigene32781_HaMG	165.44	5	Y	Buccalin precursor [ <i>Aplysia californica</i> - NP_001191649.1]
CL5959.Contig2_HaMG	117.27	5	Y	Non-neuronal cytoplasmic intermediate filament protein [ <i>Helix aspersa</i> - CAA39416.1]
CL81.Contig1_HaMG	109.71	5	Y	Actin [ <i>Biomphalaria pfeifferi</i> -AAK68712.1]
Unigene1286_HaMG	105.25	4	Y	Protein disulfide-isomerase A3 [ <i>Crassostrea gigas</i> ]
Unigene25454_HaMG	97.57	2	Y	Transgelin-2 [ <i>Crassostrea gigas</i> ]
CL6679.Contig3_HaMG	90.99	4	Y	Calreticulin [ <i>Aplysia californica</i> - AAB24569.1]
CL321.Contig1_HaMG	87.76	4	Y	78 kDa glucose-regulated protein precursor [ <i>Aplysia californica</i> - Q16956.1]
Unigene31995_HaMG	80.25	2	Y	Non-neuronal cytoplasmic intermediate filament protein [ <i>Helix aspersa</i> - CAA39416.1]
CL1651.Contig1_HaMG	79.92	3	Y	glucose-regulated protein 94 [ <i>Crassostrea gigas</i> ]
Unigene33640_HaMG	77.48	2	Y	26S proteasome non-ATPase regulatory subunit 14 [ <i>Crassostrea gigas</i> - EKC35644.1]
CL3127.Contig1_HaMG	73.1	2	Y	cholinesterase-like [ <i>Ailuropodamelanoleuca</i> ]
Unigene33436_HaMG	61.34	1	Y	actin I [ <i>Sepia officinalis</i> ]
Unigene28810_HaMG	60.62	1	Y	putative polyadenylate-binding protein 1 [ <i>Haliotis diversicolor</i> ]
CL378.Contig1_HaMG	60.62	1	Y	Polyadenylate-binding protein 4 [ <i>Crassostrea gigas</i> ]
CL4509.Contig2_HaMG	57.69	2	N	elongation factor-1 delta b isoform 1 [ <i>Daniorerio</i> - CAI21006.1]
CL4509.Contig1_HaMG	57.69	2	N	elongation factor 1-beta-like [ <i>Oreochromis niloticus</i> ]
Unigene31980_HaMG	55.7	1	N	nucleoside diphosphate kinase B [ <i>Haliotis discus discus</i> ]
CL6817.Contig1_HaMG	52.37	2	Y	arginine kinase [ <i>Biomphalaria glabrata</i> - ADH59421]
Unigene34625_HaMG	51.32	1	N	fructose-biphosphatealdolase [ <i>Biomphalaria glabrata</i> ]
CL2802.Contig2_HaMG	50.47	1	N	myosin regulatory light chain [ <i>Fasciola hepatica</i> ]
Unigene32295_HaMG	50.1	1	N	eukaryotic translation elongation factor 1 epsilon-1 [ <i>Mus musculus</i> - Q9D1M4.1]
CL5683.Contig2_HaMG	48.71	1	Y	FMRF-amide neuropeptide [ <i>Crassostrea gigas</i> - EKC40093.1]
Unigene32511_HaMG	48.71	1	Y	PREDICTED: myomodulin neuropeptides 1-like [ <i>Aplysia californica</i> ]
Unigene33563_HaMG	48.08	1	N	Endoplasmic reticulum protein ERp29 [ <i>Crassostrea gigas</i> ]
Unigene26477_HaMG	47.94	1	Y	PREDICTED: endoplasmic reticulum resident protein 44-like [ <i>Aplysia californica</i> - XP_005108389]
CL5172.Contig2_HaMG	47.89	1	N	heat shock protein cognate 5 partial [ <i>Ptyctima</i> gen. sp. - AD1287]
CL7404.Contig6_HaMG	47.89	1	N	HSP70 [ <i>Mytilus galloprovincialis</i> ]
CL6230.Contig1_HaMG	47.89	1	N	heat-shock protein 70 [ <i>Biomphalaria glabrata</i> ]
CL6830.Contig1_HaMG	47.89	1	N	heat shock cognate protein 70 [ <i>Haliotis diversicolor</i> ]
Unigene2750_HaMG	47.74	1	N	PREDICTED: muscle M-line assembly protein unc-89-like, partial [ <i>Aplysia californica</i> - XP_005111742]
Unigene20663_HaMG	47.19	1	N	PREDICTED: calcium-regulated heat stable protein 1-like [ <i>Aplysia californica</i> - XP_005098052]
Unigene24727_HaMG	46.69	1	N	putative ribosomal protein [ <i>Arion lusitanicus</i> ]
Unigene32909_HaMG	46.06	2	N	peptidylprolyl cis-trans isomerase B [ <i>Conus novaehollandiae</i> ]
CL3377.Contig1_HaMG	45.09	1	N	Heterogeneous nuclear ribonucleoprotein L [ <i>Crassostrea gigas</i> ]
Unigene4188_HaMG	44.55	1	Y	FFamide [ <i>Crassostrea gigas</i> - CU987867]
Unigene8400_HaMG	43.62	1	Y	ribosomal protein rpl38 [ <i>Lineus viridis</i> ]
Unigene34590_HaMG	42.73	1	Y	predicted protein [ <i>Nematostellavectensis</i> - EDO32930.1]
Unigene26134_HaMG	42.58	1	N	Chain A Placopecten Striated Muscle Myosin
Unigene19980_HaMG	42.26	1	N	PREDICTED: 10 kDa heat shock protein, mitochondrial-like [ <i>Aplysia californica</i> - XP_005097388]
Unigene34509_HaMG	42.22	1	N	PREDICTED: phosphoenolpyruvatephosphomutase-like [ <i>Aplysia californica</i> - XP_005106175]
Unigene32934_HaMG	40.26	2	Y	FVRamide neuropeptide precursor [ <i>Platynereis dumerilii</i> ]

Unigene33965_HaMG	40.24	1	N	ribosomal protein L10a [ <i>Littorinasaxatilis</i> ]
CL5866.Contig1_HaMG	38.71	1	N	arginase type I-like protein [ <i>Hyriopsisiscumingii</i> ]
CL331.Contig3_HaMG	35.81	1	N	acetylcholinesterase [ <i>Rhipicephalusappendiculatus</i> ]
CL331.Contig2_HaMG	35.81	1	N	Cholinesterase [ <i>Crassostreagigas</i> ]
CL331.Contig1_HaMG	35.81	1	N	acetylcholinesterase putative [ <i>Ixodesscapularis</i> - EEC16520.1]
Unigene24585_HaMG	35.49	1	Y	Myosin-VI [ <i>Crassostreagigas</i> ]
CL2369.Contig2_HaMG	34.94	1	N	Calumenin-B [ <i>Crassostreagigas</i> ]
CL1207.Contig1_HaMG	34.03	1	N	PREDICTED: glutaryl-CoA dehydrogenase, mitochondrial-like [ <i>Aplysia californica</i> - XP_005098406]
CL3253.Contig1_HaMG	32.47	1	N	paramyosin [ <i>Haliotis discus discus</i> - BAJ61596]
CL5457.Contig2_HaMG	29.71	1	N	Matrix metalloproteinase-21 [ <i>Crassostreagigas</i> ]
CL6000.Contig1_HaMG	27.02	1	N	PREDICTED: ATP synthase subunit beta, mitochondrial-like [ <i>Aplysia californica</i> - XP_005107892]
Unigene35379_HaMG	25.56	1	N	Enolase-phosphatase E1 [ <i>Crassostreagigas</i> ]
CL1213.Contig1_HaMG	24.27	1	N	DnaJ-like protein subfamily B member 11 [ <i>Crassostreagigas</i> ]
CL6818.Contig1_HaMG	23.68	1	N	Protein disulfide-isomerase A6 [ <i>Crassostreagigas</i> ]
Unigene32024_HaMG	21.1	1	Y	Protein [ <i>Xenopus (Silurana) tropicalis</i> - LOC100145586]
Unigene33735_HaMG	18.74	1	N	Sphingomyelin phosphodiesterase 4 [ <i>Crassostreagigas</i> ]

**Table S3. Summary of proteins with no BLASTp match in RP-HPLC fraction 21.**

Accession	-10lgP	#Peptides	PTM	Description
CL157.Contig1_HaMG	120.21	7	Y	Hypothetical
Unigene437_HaMG	120.21	7	Y	Hypothetical
CL157.Contig2_HaMG	120.21	7	Y	Hypothetical
CL1788.Contig3_HaMG	107.23	2	Y	Hypothetical
CL1788.Contig1_HaMG	102.55	2	Y	Hypothetical
CL1788.Contig2_HaMG	102.55	2	Y	Hypothetical
Unigene16914_HaMG	96.59	3	N	Hypothetical
CL157.Contig3_HaMG	89.75	4	Y	Hypothetical
Unigene14067_HaMG	77.67	1	N	Hypothetical
CL389.Contig1_HaMG	76.46	1	Y	Hypothetical
Unigene11859_HaMG	61.7	2	Y	Hypothetical
CL6297.Contig2_HaMG	60.42	1	Y	Hypothetical
CL7284.Contig12_HaMG	53.87	2	N	Hypothetical
CL1256.Contig2_HaMG	50.95	1	N	Hypothetical
CL4331.Contig1_HaMG	50.26	1	N	Hypothetical
CL5683.Contig1_HaMG	48.71	1	Y	Hypothetical
Unigene1414_HaMG	48.71	1	Y	Hypothetical
CL157.Contig11_HaMG	47.75	1	N	Hypothetical
CL6105.Contig1_HaMG	47.21	1	N	Hypothetical
CL5930.Contig1_HaMG	46.24	1	N	hypothetical protein [ <i>Amblyommamaculatum</i> ]
CL5630.Contig1_HaMG	45.11	1	N	hypothetical protein [ <i>Crassostreagigas</i> - CGI_10017178]
Unigene4188_HaMG	44.55	1	Y	hypothetical protein [ <i>Crassostreagigas</i> - CGI_10013719]
Unigene35445_HaMG	36.64	1	N	Hypothetical
CL1643.Contig1_HaMG	34.68	1	N	Hypothetical
Unigene25376_HaMG	32.04	1	Y	Hypothetical
Unigene19464_HaMG	26.96	1	Y	Hypothetical
Unigene20959_HaMG	26.62	1	N	Hypothetical
CL1810.Contig2_HaMG	25.91	1	N	Hypothetical
Unigene42640_HaMG	25.47	1	Y	Hypothetical
Unigene587_HaMG	24.93	1	N	Hypothetical
Unigene35738_HaMG	24.3	1	N	Hypothetical
Unigene33522_HaMG	24.25	1	N	Hypothetical
Unigene12754_HaMG	24.01	1	Y	Hypothetical
Unigene43128_HaMG	23.67	1	Y	Hypothetical
CL4417.Contig3_HaMG	23.25	1	N	Hypothetical
CL6997.Contig1_HaMG	22.89	1	Y	Hypothetical
CL6518.Contig1_HaMG	22.78	1	N	Hypothetical
Unigene11807_HaMG	22.46	1	Y	Hypothetical
Unigene41971_HaMG	22.11	1	N	Hypothetical
Unigene35343_HaMG	21.91	1	Y	Hypothetical
Unigene26746_HaMG	21.5	1	N	Hypothetical
Unigene20095_HaMG	21.35	1	Y	Hypothetical
Unigene41299_HaMG	21.29	1	N	Hypothetical
CL449.Contig3_HaMG	21.2	1	Y	Hypothetical
Unigene33873_HaMG	21.14	1	N	Hypothetical
Unigene17097_HaMG	21.1	1	N	Hypothetical

CL1651.Contig2_HaMG	21	1	N	Hypothetical
Unigene25903_HaMG	20.76	1	N	Hypothetical
Unigene32393_HaMG	20.49	1	Y	Hypothetical
Unigene38751_HaMG	20.48	1	N	Hypothetical
Unigene25632_HaMG	20.04	1	N	Hypothetical
Unigene35028_HaMG	18.69	1	N	Hypothetical
Unigene8978_HaMG	18.42	1	Y	Hypothetical
Unigene32284_HaMG	17.92	1	N	Hypothetical
Unigene25385_HaMG	17.17	1	N	Hypothetical

**File S1 – List of *Cornu aspersum* mucous gland precursor proteins derived from RP-HPLC fraction 21.**

>Unigene32781\_HaMG -3 252 959  
MLRKTPLQQLLLLALVLTNLQISASDVSDTADSPDAFNDITASKDAADIEHAVLSAREDQDSEDPDDVEKRKLD SYGFY G  
GIGKRRVDRFLFAEGIGKRQLDPFSFAGHLWKRSINNTGGSGNRRLDKNGFAGQIGKRSLDPFGFAGQLGKRGLDKS  
GSA  
GAIEKRQLDTNGLAGGIGKRQANIKIDIYGFAGGIGKRSEEDGFKYDDIDDVEAESENDRHVDRSAQLDKQVHVI  
>CL5959.Contig2\_HaMG +1 256 1983  
MTSKISTTYEEGRQSKIQPRAFVITRSGPSSKSSSFARQSYASSRQSITPGVYQQLSSSGITDFRGTREKEKREMQN  
L  
NERLASYIEKVHFLDAQVKKLEAENEALRNRKSESLQPIRDAYENELAQARKVIDELSSSTKGVSEAKVAGLQDEIASLRE  
LIVTYENQSKDYRKKIESLGNQIGEYEGELHTLRIRCGSLEDENAKVRELLDKIQEQNRRLRADLDTETAHIEADCLAQ  
TKTEEAIFYKDLLDQLELLKPEPIQIKGMDYAEFWKSELSKCVREIQSAYDEKIDMIQQDTEAKYSAQLNSLRSGNVKD  
G  
MQLQHVQEEVKLRTQAGEKNAMYAELAAKFASLQAERDSIGRQCSELERELEELRIKYNQDIGDLSNELSAVLAQLQI  
L  
TDAKITMELEIACYRKLLEGEESRVGLRSLVEQAIGVQGRGTASLKDTIQQSTSSGSMTIQRSSKGPIAFNSVDQSGSNI  
VIENTTSGARAKTQSLRGWRIDKTVAGRVAASIQLKDYEIPPNTKYTIWAKGAKDRATADNEQIADIFSLGVGSCTWTIV  
DEAGNEKATLVAKFSG  
>CL81.Contig1\_HaMG -3 120 1247  
MCDDDVAALVVDNGSGMCKAGFAGDDAPRAVFPVIVGRPRHQGVMVGMGQKDSYVGDEAQSKRGILTLYPIEHGIV  
TNW  
DDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMTQIMFETFNTPAMYVAIQAVLSLYASGRRTTGIVLDSDG  
V  
THTVPIYEGYALPHAIMRLDLAGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAATSSSLEKS  
YELPDGQVITIGNERFRCPEAMFQPSFLGMESAGIHETTYNSIMKCDVDIRKDLIANIVLSGGSTMFPPIADRMQKEISA  
LAPATMKIKIIPPERKYSVWIGGSILASLSTFQQMWISKQEYDESGPSIVHRKCF  
>Unigene1286\_HaMG -3 114 1598  
MKTYILYLCVVLYSAVQAADDVIVLTDSSFASEVAKHPVVLVEFYAPWCGHCKRLAPEYEKAATILSRSDPPVPLAKVDC  
TVETSVCSKYGVSGYPTLKIFKNGEFSKEYEGPRETDGIVKVMNREAGPVSKLTTKEEAQNFINKDTVGVIGFFDAED  
N  
AEAKSFLKVADQLPNIRFAHTTSDEVKEEFKQTEQGVVLYRPKVLQSKFEDAEVVRTTETGNKLSILESEALGLCGERT  
A  
GNADTFPKPLVVAYFKVDFRKSQKQTNWRNRLMKVVKGLKDDKLTVSCAISDIDELGRELEEVGIEDRSGDKPFVVA  
YD  
SSNKKYKMAAVFSVETFEFVDRDFLGGKLEPHIKSEPIPEPNDDAVKVVVAKNFDEIVNNEKDVLIEFYAPWCGHCKS  
L  
APKYEELAKKLEGETDIVIAKMDATANDVPSQYEVIRGFPTIYFAPKGSKSPRKYESGREVDDFIKYLAKESTDGLSGY  
D  
RDGKKKKEKKKTEL  
>Unigene25454\_HaMG +1 139 633  
MADTRAKKSGLGADIERKLEGRYDEEEAAGTMNAIVVWINAVLQGEHEPIANQSHKELHRALRDGVLLCKLINKLLASE  
G  
KGKVSFQKKVASPFVAMGNIESFNKGCLEYGLSKEFLFQSGDLWEGRKGPFLNVINGIHSGLFLANTKGFQPTYTGQQ  
TK  
YVDNE  
>CL6679.Contig3\_HaMG +2 212 1330  
MTAAALLLLLLGVACADVVFKEEFNDGDGWLRSRIVESKHKSDIGKFTRTAGKFGDAEKDQGIQTSQDARFFGLSARF  
DK  
FSNEDKTLIIQFTVKHEQNIDCGGGYVKVFPSQLDQADMHGETPYNIMFGPDICGSATKKVHVIFNYKGNLLIKKDIRC  
KDDVYTHLYTLIVRPDNTYEVKIDNEKVESGTLEEDWDFLPPKTIPDPDAKPPDDWDEMDGEWEPPEMIDNPEYKGE  
WKP  
KQIDNPDYKGRWVHPEIDNPEYVADAKLYKYDDFGAIGFDLWQVKAGTIFDNILITDDVDYAEFFGQETWGKTKDAEKK  
M  
KEAQDEDERKAREEEEEKRRKEEEDAAGDEEAEADQEEEEKDDDEEDVAKDEL  
>CL321.Contig1\_HaMG +1 139 2127  
MEKLVAPYLLLLALVISTSMVCAKDDDDGEEKKSDVGTVIGIDLGTTYSCVGVFKNGRVDIIANDQGNRITPSYVAFTPD  
GERLIGDAAKNQLTSNPENTIFDVKRLIGRTWDDKSVQQDIKYPFKVIEKSSKPYVQVDTSEGDKTFAPEEISAMVLSK  
MREIAEAYLGKKIVNAVVTVPAYFNDAQRQATKDAGTIAGLNVMRIINEPTAAAIAYGLDKKEGEKNILVFDLGGGTFDV  
SLLTIDNGVFEVIATNGDTHLGGEDFDQRVMDHFILYKKKKGDVVRKDNRAVQKLRRREVEKAKRALSSSHNARLEIEG  
F

FENEDFSETLTRAKFEELNMDLFRNTMKPVAKVLEDADLKKDDVHEIVLVGGSTRIPKVQQLVKEFFNGKEPSRGINPD  
E  
AVAYGAAVQAGVLGGEEDTGDVLVLLDVNPLTLGIETVGGVMTKLIPRNTVIPTKKSQIFSTAADNQPTVTIQVYEGERSM  
TKDNHLLGKFDLTGIPPAPRGVPQIEVTFEIDVNGILRVTAEDKGTGKNHITIQNDQNRLSPEDIDRMINDAEKYADED  
KKIKERVDKNELESYAYSLSKNQIGDKEKLGAKLSEDDKEKITEAVDGAIKWLESTPDAEVDEFKEKKTELEGVVPIMT  
KLYEQSGGAPPPTDDSDSDKDEL  
>Unigene31995\_HaMG +1 256 1614  
MTSKISTTYEEGRQSKIQPRAFVITRSGPSSKSSSFARQSYASSRQSITPGVYQQLSSSGITDFRGTREKEKREMQN  
L  
NERLASYIEKVHFLDAQVKKLEAENEALRNRKSESLQPIRDAYENELAQARKVIDELSSTKGVSEAKVAGLQDEIASLRE  
LIVTYENQSKDYRKKIESLGNQIGEYEGELHLRIRCGSLEDENAKVRELLDKIQEQNRRLRADLDTETAHIEADCLAQ  
TKTEEAIFYKDLLDQLELLKPEPIQIKGMDYAEFWKSELKCVREIQSAYDEKIDMIQQDTEAKYSAQLNSLRSGNVKD  
G  
MQLQHVQEEVKKLRTQAGEKNAMYAELAAKFASLQAERDSIGRQCSELERELEELRIKYNQDIGDLSNELSAVLAQLQI  
L  
TDAKITMELEIACYRKLLEGEESRVGLRSLVEQAIGVQGRGTASLKDTIQSS  
>CL1651.Contig1\_HaMG-2 80 2479  
MAKMKQYLIIGVACLLFAVAAAADDDDDGDLTDDDIGKSRDASRTDDEAVQREEEAISLDGLSVAQMKELREKSEKF  
A  
FQAEVNRMMKLIINSLYKNKEIFLRELISNASDALDKIRFISLTDNSALAATEELTIKIKADKDNKVLHITDTGIGMTKD  
DLIKNLGTIAKSGTSDFLAKLGDANSATETSDLIGQFVGFYSAFLVADRIVVTSKHNDQYIWESDAENFSIIKDPRG  
NTLGRGTTVSLQLKEEATDYLEEATLKNLVKKYSQFINFNIYLWTSQEQQVEEPIEDDDAEKAEKESDDDDDEEEAKVE  
E  
EEDGDKPKTKKVTKTVWDWELMNSVKPIWTRKSSDVSDEYNAFYKSISKESEPMAKIHFTAEGEVTFKSILFVPQTS  
P  
YDTSFNYGKKVDHIKMYVRRVFITDDFEDMMPKYLNFIRGVVDSDDLPLNVSRETQQHKLKVIKVKLVRKALDMIKKI  
DKETYDKFWKEYSTNIKLGVIEDTSNRTRLAKLLKFYSSNSQTEQTFADYLERMKDKQEAIFYAGTSREEVENSFPVE  
RLLKKGVEVLYLTPVDEYCIQSLPEFEGKFKQNIKEGVNLDTSEKAKERKEALEKEFKPLTDWLKDEALKDNIEKATI  
SERLTPCALVASQYGWGSGNMRIMKSQAYAKQGDASSQFYASQKKTLEINAKHPLVRKLKELIADPSDENAKDLA  
KV  
LFDATLRSQYQVKDTLDFAKRIEMLRSTVDIDVNEVIEEPEEEEEVAADNTNDDAAEEDDEEVADEDKDSDESAKDE  
L  
>Unigene33640\_HaMG -2 65 997  
MERLLSLGARVPGLGQGSPTDAPVVDTAEQVYISSLALLKMLKHGRAGVPMVEMGLMLGEFVDDYTVRVIDVFAMP  
QSG  
TGVSVEAVDPVFQAKMLDMLRQTGRPEMVVGVYHSHPGFGCWLSGVDINTQQSFEALSERAVAVVVDPIQSVKGV  
VIDA  
FRLINPNMMVLGQEPRQTTSNLGHLNKPISQALIHGLNRHYYSIAINYRKNELEQKMLLNLHKKSWVDGLQLQNYSEHC  
S  
LNETTVGEMLELAKNYHKALEEEETMTPEQLAIKHVGKQDPKRHLEEAVGVVMTKNIVQCLGAMLHTVVFK  
>CL3127.Contig1\_HaMG+3 57 1742  
MCSMAGAAYLVLLTAAVLIGCVHSQDTCRQTRTNLGPILGKINSVNVAGRLVKVETFWGIPYVKPPIGNLRFQRPIPEM  
WTEVRRWCWVPPKACLQEVDNDIIQSAEDMSEDCLYINVWRPLCNTTNMATIVWIHGGSLRSGSSAYPLYNGEYLAAR  
ECV  
IVASMNYRLGALGFFYQGGTNFSFFTNLGFMVDEQALACIWWNVEDFGGSRKITVAGSGAGAYLASLLLLVNIPRVNF  
T  
SAILQSGGPLMKDLTQTLDDAKSAGSQLARDLQCPYTNDTELMECLSKVDGVVLTAKAYMKMAAGTQHKLVTMGGIFQL  
DK  
MPLEYLQNGEIKNLPILVGSNRDEGTEILMKIFPSDFNTTQPPPEFSDQKLDDALTQILGRFFKPPHLPDLITRLKGI  
DYEMRNETNPTVTLLSAVIRDLTYTCPLYAFLEARANHLNSQNVYMYSFNNVPVSTQLPAWVGATHGTELQYLFGTS  
NT  
YMPADLSPSETALSLTMQKMWANFAKFGNPTDNTVGTQWPPFTYRTSERMMVFTENAQLNVAAFNNTVCTFWNSI  
PPEQ  
RV  
>Unigene33436\_HaMG -3 63 620  
MKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSYELPDGQVITIGNERFRCPESLFQPSFLGM  
ESAGIHETTYNSIMKCDVDIRKDLANTVLSGGTTFMFGIADRMQKEITSLAPSTMKIKIIPPERKYSVWIGGSILASL  
STFQQMWISKQEYDESGPSIVHRKCF  
>Unigene28810\_HaMG -2 245 928  
MNPATSYPMASLYVGDHPDVTEAMLFEKFSTCGPVLSIRVCRDMITRRSLGYAYVNFQQPHDAERAMDTMNFDTIKG  
RP  
IRIMWSQRDPSLRKSGVGNVFIKNLDSIDNKALYDTFSAFGNILSCKIACDEHGSKGYGFVHFETEEAARQSIEKVNGM

LLNGKKVYVGRFIPRKERIMLMGDKHRKFNNVYIKNFADELDDHRELFEQYGKIISAKVMYDLSGK  
>CL378.Contig1\_HaMG -2 269 2131  
MNPATSYPMASLYVGDLPDVTEAMLFKFKSTCGPVLRSIRVCRDMITRRSLGYAYVNFQQPHDAERAMDTMNFDTIKG  
RP  
IRIMWSQRDPPLSRKSGVGNVFIKNDKSIDNKALYDTFSAFGNILSCKIACDEHGSKGYGFVHFETEEAARQSIEKVNGM  
LLNGKKVYVGRFIPRKERIMLMGDKQRKYNNVYIKNFGDELDDAKLRELFEHYGKIISAKVMYDNGKSRGFGFVSFEE  
P  
EAAEKAVSMLNGSELGGKQLYCGRAQKKTERQAELKEKFEKIKLERINRYQGVNLYVKNLDDGVDDERLRKEFSQFGT  
IT  
SAKVMGEGGRSRGFGFVCFSSPEEATKAVTEMNGRIIVAKPLYVALAQRKEDRRAHLASQHIQRLASLRQPNIQYGS  
FN  
AGGAGYFVPTMPQAQRGFFTPANVPTIRQARTPWAPGVRPNQAAGFQTMPNQIRTRPPAAQAQMRPTAINSRPITGQ  
PIA  
AARGVPIAGVARPPGVAAMADPGSRPAPYQPGPRQMPGQPGRPMPQPQQSVVIAGQDPLTASMLAAAPPQEQQM  
LGERL  
FPLIQHMYPELAGKITGMILLEIDNSELLHMLESQESLKLKVDEAVAVLQAHQAKESATKKE  
>CL4509.Contig2\_HaMG -3 3 497  
TYGPAGAPASDASPAAPKAPAAAAPKAPAAAAPKAAAGDDDDDELDFGSDDEETKAIKAKRLEEYATKKAKKPVLI  
AKS  
NIILDVKPWEDTTDMDEMERCVRVQTDGLLWGASKFVPVGYGIRKQISCVVEDDKVSTDFLEEEITKFEDLVQSV  
DIA  
AFNKV  
>CL4509.Contig1\_HaMG -2 83 784  
MGFGDLKSQAGLQVLNLFLEDRSYIDGYVPSQADIAVFLAVSSPPASLPHALRWYNQVASYESQKDSLPGVRSAD  
T  
YGPAGAPASDASPAAPKAPAAAAPKAAAGDDDDDELDFGSDDEETKAIKAKRLEEYATKKAKKPVLI  
AKS  
NIILDVKPWEDTTDMDEMERCVRVQTDGLLWGASKFVPVGYGIRKQISCVVEDDKVSTDFLEEEITKFEDLVQSV  
DIA  
AFNKV  
>Unigene31980\_HaMG -1 19 570  
MTRSPSFLSGCIRMVATLLSLLSLFGKNMADPRAERTFLAVKPDGVKRGVLGKIVARFEERGYKVFAGKLVTPPL  
LLR  
S  
HYKDLSDRPFPPGLIKYMSSGPVFAMVFEGTGIVKMARMMLGATNPQDSTPGTIRGDYCIEVGRNVCHGSDSV  
ETAER  
EI  
ALWFTQEELAEYVVPNKELIFESN  
>CL6817.Contig1\_HaMG +3 189 1256  
MAAAEAQKLYELLKADAKCKSLLSKHLTQALVDQLKDKKTKFGGTLADCIRSGCENTDSGVGVYASDPDAYTVFAP  
LL  
DA  
VIKDYHKVSELNHPNPDFGDVEKLGFGDLDPGKLVSTRVVRVGRSHDSYGFPPVLTKEDRVEMERRTVEAFKLD  
GE  
LK  
GTYHPLTGMSKETQKQLTEDHFLFNDSRFLRAAGGYNDWPTGRGIFNDNKTFLVWVNEEDHLRFISMQKGGN  
LQE  
VYV  
RLVKAIKALESSGLSFAKRQGLGYLTFPCPSNLGTTLRASVHIKIPKLAASPDFKAFCDKYNIQARGIHGEHT  
ESVGGVYD  
ISNKRRRLGLTEIFAIQEMRIGVEAVIAEERRLLGGGN  
>Unigene34625\_HaMG +3 48 1148  
MAGSSTRFAQYLTPQENELRGIANTIVGKYRGILAADESVAFIGRFDITIGLTNTEENRRKYRELLFTT  
DNSIADNIG  
VILFDETFRQSASDGTTPFVKVLQNKGIIPGIKVDKGVVPLAGTINECTTQGLDGLAERCAEYKKGCGF  
AKWRCVLKISE  
HTPSYQAMIENANVLARYASICQQNGLVPIVEPEVLDPDGTDLQTAQRVTEEVLAQYKALADHHVFLE  
GTLLKPNMVT  
A  
GQGCSKKYTSQDIARATVEALNRTVPPAVAGITFLSGGQSEEDATLNLNAINQYPGRKPWPLTFSYGRAL  
QASVIKAW  
GG  
KDGNVKAAQEELFKRARSNGLAAAGEYQGGGAAGAAGGQSLFVANHSY  
>CL2802.Contig2\_HaMG +1 97 579  
MAEEKHARHTTNILASFRQAQVQEYKEAFTMIDQNRDGIIDIEDLKDMYSNLGRIPPDSELNEMLKEAPG  
PLNFTMFLNL  
FGEKLSGTDPEDTIRQAFSMFDEAGKGFPIPEEYKDLLANMGDNFTEDEIKQTWKEAPLEKGFYDNTFV  
GILKGGKEED  
A  
Q  
>Unigene32295\_HaMG -3 195 515  
MQVEQWLEYTLYIDNLQDGESERAVLKELNSYLQDKVYFVGHKLTVADILLYHSLHRVFKHLSFQDKEK  
FMNLSRWFD  
NV  
QHRRRLRQSLTLLPFVRSPLYSGINSH  
>CL5683.Contig2\_HaMG +1 265 792  
MTSSKYAMLSPCLGLLLASLVAASSATESLAGNPSNNAKITSSADQSDSPNRSKRMTDFALLGQDAYADKQ  
ADIPELS  
E  
LNAAEAESGKYLDEDDDHLDKAEESDPWKNMFIRIGRDPHSADQIYSNAAETDKRASQFMRIGKSPSSFV  
RIGKSEN

GN  
TENVDKRASSFVRIGK  
>Unigene32511\_HaMG +2 377 904  
MTSSKYAMLSPCLGLLLASLVAASSATESLAGNPSNNAKITSSADQSDSPNRSKRMTDFALLGQDAYADKQADIPELS  
E  
LNAAEAESGKYLDEDDHLDKAEESDPWKNMFIRIGRGPDSADQIYSNAAETDKRASQFMRIGKSPSSSFVRIGKSEN  
GN  
TENVDKRASSFVRIGK  
>Unigene33563\_HaMG +3 12 770  
MSAAGTRLGETLSLVLLSVVYFVSAEVIQGSVTLNTAVFAKVIDKHKAVLVKFEDETYPYGDKQDAFKEVAKSSIQQPDLL  
VAEVPVSEYGDKENLDLAEKYQASKDSFPAYRFLFLNGNTEKPIAFTGNTESADAIKNFVTQESGFWLGRPGCLEEFDKL  
V  
KEFLTSAKDKRGAVVEKAEQAAQRLTQDADKASADVVIKTMKKVLEKGDSEFIKSEVSRVEKLRTGKISDKKKEQLKERL  
N  
ILATFQIHVRDEL  
>Unigene26477\_HaMG +3 84 1352  
MFSVLLTSPFVLAVITLFSHVSGQVIALNEHNVDSVLAGSDIAIVNFYADWCRFSQILAPVFEEAAKKVHDEFVPGKI  
VFAKVDCDKETNIGTRFHISKYPTLKVFRGGVMVKKEYRGQRSAEALSSYMQEAKDPVHEVAHMDNLDLDPKRRHI  
IG  
CFDSKVSDNYRTFMRDANILRDDCFHAAIGPVCQADRAAGDTVSRFPNTIHQDVLYTGSLTDFEALHLWSASKCVPI  
V  
REITFENAEELTEEGLPFLILFHPDDSASVEFFNEEVNKQLLQEKANINFLVADGIKFAHPLHHLGKSTNDLPLLAIDS  
FRHMYLFPHDVKKDLATPGLLRQFVEDLHSGKLRHREFHHGPDPTTAPEPHPPVQDGSSEKNIPRDTSDPEKPPSQQKI  
TS  
PPESSFRKLAPSRNRYTILRDEL  
>CL5172.Contig2\_HaMG +2 2 322  
FTLVGIPPAPRGVPQIEVTFDIDANGIVNVSARDKGTGKEQQIIVQSSGGLSKDEIENMVRNAELHAEEDKKKKEMIEVV  
NQADGIIHDTESKMEEFKDQLPKEECD  
>CL7404.Contig6\_HaMG -2 83 799  
MTALIKRNTTIPTKQTQTFTTYSNQPGLVLIQVYEGERAMTKDNLLGKFELTGIPPAPRGVPQIEVTFDIDANGILNVS  
ASDKSTGKENKITITNDKGRLSKEEIERMVNEAERYKNEDEKQKNRIAAKNALESYCFHMKSTVEDEKLKDKISADDDK  
T  
ITDKCSEVITWLDANQLADQEEFEHKQKEIEGICNPIITKLYQGMGGAAGGGMPNFGTAGAGAGGAGSAGGGPTIEEV  
D  
>CL6230.Contig1\_HaMG -2 224 991  
MGDTSETIKDVLLVDVAPLSLGIETAGGVM TALVKRGTTIPTKTSQIFTTYSNQPGLVLIQVYEGERTMTKDNLLGKFH  
LTGIPPAPRGVPQIEVTFDIDANGILNVTAMDKSTGKSSNIVIKNERGRLSQAIEDRMLAEAERYKEEDEKQRVVRTARN  
QLENYVFSVKQAVEAAGDKLPSSDKNTVLQACEETMKWLDNNSLAEKEEFEDKLKEIQRISSSCMAKLHSQQQSGETP  
NY  
GQASGGHSGPTVEEVD  
>CL6830.Contig1\_HaMG -1 82 2025  
MSKAPAVGIDLGTTYSCVGVFQHGKVEIANDQGNRTTPSYVAFTDNERLIGDAAKNQVAMNPENTVFDKRLIGRRFD  
D  
PTVASDMKHWPFEVINDAGKPKIRVEYKGEKKOFFPEEISSMVLTKMKETAAYLGKTVTDAVVTPAYFNDSQRQAT  
KD  
AGTIAGLNVLRRIINEPTAAAIAYGLDKKVGGERNVLIFDLGGGTFDVSILTIEDGIFEVKSTAGDTHLGGEDFDNRMVNH  
FIQDFKRKHKKDISDNKRAVRRRLRTACERAKRTLSSSTQASIEIDSLFEGIDFYTSITRARFEELNADLFRFTQEPVEKS  
LRDAKLDKSQINEIVLVGGSTRIPKIQKLLQDFFNGKELNKSINPDEAVAYGAAVQAAIHLHGDKSEEVQDLLLLLDVTPLS  
LGIETAGGVM TALIKRNTTIPTKQTQTFTTYSNQPGLVLIQVYEGERAMTKDNLLGKFELTGIPPAPRGVPQIEVTFDI  
DANGILNVSAAADKSTGKENKITITNDKGRLSKEEIERMVNDAEKYKNEDEKQKTRISAKNALESYCFHMKSTVEDEKLK  
D  
KLNADDRKIITDKCSEVISWLDGNQMAEQEEFEHKQKEIEAVCNPIISKLYQGSAGGMPNFSAGGAGAGPGAAGAGGS  
SS  
GPTIEEVD  
>Unigene2750\_HaMG -1 1 1191  
TSELQAEANLTSKPASELQAEATLTSKPEDSSEAADAQFHLEKPSGQSEQRKPKDQGTGTGLRRELQPLVQAEFFPEA  
QAE  
TAPVSKSTSSLEAASVPAKTGLRKTSLAEAEQPLHSQAEARFSQVSKPLETAAEVSVQVSKPLEAAAQEVVRSKVVETPSV  
T  
EVTEREKKTRPTETKPRGEEEEKYQVEDRLEKQEQSEEFKEEKLPREERKPEQTQPQKLSKEETREEKHKEDTHRPL  
APR  
IEEMMDDCVTELGELARFDCRVSAYPDPEITWYKDGKNTVPSAKYELLNFHDDIFSLLIKVKVDKADSGRYSCVAKNEYG

E  
AKSEALLNMGLKEKIADAHVAPSFVTKFYDIEVVEGVPVEFVCMIDGQDPVVTWLLNGQEVENNSEMLTRRQEDT  
>Unigene20663\_HaMG +2 116 547  
MSQPYDIPETQAINSAQISGSPGRAFLIPSPVPTRRNRKTKSMSERAAEGPTHKGTVKFFCRQRGHGFIHPDDGHGEDV  
FL  
HISDIEGEWVPKEGDQVYKTVPIPPKNEKLQAVHAVITHLKPGLVHERWDSVPPTSPEKFSPO  
>Unigene24727\_HaMG -1 46 522  
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLLRGGGKKR  
KKKNYTPKKNKHKKKKVLAFLVLYKVDENGGITRLRRECPNKCCGAGVFMASHFDRQYCGKCCLTYVFNKPGEEV  
DS  
>Unigene32909\_HaMG +1 145 771  
MLLAAAVLLLSFSGLALSEKAKGPKVTDQVFFDIEIGGKPAGRIIIIGLFGKTVPKTVENFKALAAGDRKDGDTVLAFKGS  
KFHRVIKDFMIQGGDFTRGDGTGGKSIYGDKFDENFKLKHGAGWLSMANAGKDTNGSQFFLTVKKTEWLDGRHVV  
FGK  
VVEGMSVVRKIESTPTDGRDKPKEDVVIVDSGVLEVTEPFTVEKADATE  
>CL3377.Contig1\_HaMG +3 141 1520  
MLMPKRQALVEFEDLEGAKNCVEYCQQGNAIYCGGQPAFFNYSMSQKIERSVEMDDSKPPNHILLFTVLNPQYPITV  
DV  
MNTINKPYGCVQRIVIFKNGVQAMVEFDNVESAKRARQALNGADIYSGCCTLKIEFAKPARLNVVRNDHDSWDYTNP  
AL  
GGNKLEPVARGQPLLQEPRIQSAPAPYNGPVGSGFAPGPSGPPVFGNPGYYDDGFGRPGPGYGGPGDRYGARPDR  
FGHG  
GYEVLGGQGTVLMVYGLNMEKINCDRLFNLFLYGNVIRIKFLKSKEGSAMVELADSLSVERAISNLNNAFLFGSKLQL  
SLSKQSHLQEVMPHQLPDGTMSFKDFSENRRNRFSPNAAHKNRIQNPSKVLHFFNAPPNMGEKEISELFENNYGK  
SPV  
AVKTFPSKTERSSAGLVEFENKSDGVEALVMVNHTAVNTPESKTPFIFKLCFSSMPAAST  
>Unigene4188\_HaMG -2 263 535  
MKCRFLFVVISIAALLITKSHVTAQRNIKDYGQQPLLFGRRGLNPNMNSLFFGKRGAINSEVSLRDLKTACAMILPYL  
EDADSTVDEEV  
>Unigene8400\_HaMG -1 40 249  
MPKQIQEIKDFLLTARRKDAKSVKIKKNDNVKFKVRCRYLYTLVIQDREKAEKLRQSLPPGLAVKELK  
>Unigene34590\_HaMG -2 56 439  
MKKIADPKWEPEPEVVITLTKDNFDETVNSEKMLLVEFYAPWCGHCKKLTVPYEKAAKELSKVEPPILLAKVDATKESE  
L  
AVRYNVTGYPTLKVFRSGKATDYRGERNTPYDIASYMKNQIGDGAKEI  
>Unigene26134\_HaMG +3 3 254  
SKGEEITVKLVNSAETRTVKKDDVQQMNPPEYKIDDMANMTYLNEASVLHNLKSRYAGLIYTYSGLFCVAVNPPYRRL  
P  
IYTQ  
>Unigene19980\_HaMG +3 93 395  
MAQAFKRFLPLFDRVLVERFLPELKTGGIILPEKSQGVVEATVVAIGAGARKDNGTVVAPSVKVGDKVLLPEYGGTK  
V  
VIEEKEYFLFRDGDLLGKFEN  
>Unigene34509\_HaMG -2 2 667  
MNDACSIPILVDADTGYGNFNNARRLVKLEDRGIAGAAIEDKLFPKTNSLLDGRRQPLADVKEFSLKIQAMKDTQRDP  
D  
FNIVARVEALIAGWGVDEALKRADAYLAAGATAILMHSKLKEPTDILDFLKAWKNKGPVVLVPTMYKTPMQVWRDHK  
VS  
LIIWANHSMRASVKAMQDLTAQIYKEQSTVGVENKIATVKELFRLTKDDELKTAENKYLPOV  
>Unigene32934\_HaMG +2 2 727  
PSKFVRIGRANSHFDRIKSPSNFVRIGKSGDAFDENDVNTDYAEEADNFEKRASSFVRIGKAPSSFVRIGKAPSSFVRI  
GKADQIADNLGADSIENNRASRFVRIGKSGIEDEINDAKRASSFVRIGKSGSKSTDNSDPQKRASSFVRIGKSSLSDSQ  
NSALASPADSSSKSSEFLDQIPEDNQNPVHARGSAFVRIGKIPSSAFVRIGKNNGLLIDEIESNRNTRGSQSSFVRI  
GK  
>Unigene33965\_HaMG +2 83 733  
MSKINRDSLRESIDNVLKFATDTKRRKFTESVELQIALKNYDPQDKRFAGTIKHKHIPKPKMKVCVLGDQVHCDQAKAN  
DLPHMDIEGLKKNKDKKLVKKLAKRYDAFLASESLIRQIPRVLGPGLNKAGKFPAPITHGDVLNAKVEETKATIKFQMK  
KVLCLAVCVGHVNMSSSEELYQNITLAVNFLVSLKKNWQNVRALYIKSTMGPSEKIF  
>CL5866.Contig1\_HaMG -1 1 852  
VLRAGDLAEKLEALGNTVKDYGDVAITEIPNDPPVDNAKNPKAVGAANKALSIAVAKAVKETSSALILGGDHSMGIGSIY  
GHAQVQPNLVVLWIDAHADINTPLTSTTGNMHGMPVAFVARELQKNVPKLPGFEWLKPCISVKDIAYIGLRDVGPERA  
I

LEKHGIASYSMQEVDQYGIANVVDRAIKQVDPSGNRPIHLSFDVDSMDPSVTPSTGTPVPGGLTRDASYLGEIIAATNR  
LAVLDITEVNPCLGSAEDSKLTVKSTIDVTTTTFYGRKRQGNIPA  
>CL331.Contig3\_HaMG -3 39 878  
MCWMA SPVCPMLLVA AVLIRCVHSQTSSV VISTSTGSFRGTSKNILINGRQIPVETYWGIPYALPPTGERRFKRPEPITQ  
PSTTVYNATVPPKACLQDVVNRRTIRSLMDTSEDCLYFNVWRPRCNISNMATLVHIIHGGSLRSGSSTNPLYKGENLAAG  
EC  
VLVMSFNRYRLGALGFLHVDTPGFHSNAGLYDVKHALLKIKNEVSSFSGNRDKITVFGSGAGAF LASTFVFSDDL RDSVN  
N  
VILQSGSPTMQMLYQSRTSAMTAAKRLGTDLQYLFVLSGP  
>CL331.Contig2\_HaMG -3 39 935  
MCWMA SPVCPMLLVA AVLIRCVHSQTSSV VISTSTGSFRGTSKNILINGRQIPVETYWGIPYALPPTGERRFKRPEPITQ  
PSTTVYNATVPPKACLQDVVNRRTIRSLMDTSEDCLYFNVWRPRCNISNMATLVHIIHGGSLRSGSSTNPLYKGENLAAG  
EC  
VLVMSFNRYRLGALGFLHVDTPGFHSNAGLYDVKHALLKIKNEVSSFSGNRDKITVFGSGAGAF LASTFVFSDDL RDSVN  
N  
VILQSGSPTMQMLYQSRTSAMTAAKRLGTDLQCNTDNMTLLVSCLRQICLCCPVPDEDL  
>CL331.Contig1\_HaMG -3 39 1628  
MCWMA SPVCPMLLVA AVLIRCVHSQTSSV VISTSTGSFRGTSKNILINGRQIPVETYWGIPYALPPTGERRFKRPEPITQ  
PSTTVYNATVPPKACLQDVVNRRTIRSLMDTSEDCLYFNVWRPRCNISNMATLVHIIHGGSLRSGSSTNPLYKGENLAAG  
EC  
VLVMSFNRYRLGALGFLHVDTPGFHSNAGLYDVKHALLKIKNEVSSFSGNRDKITVFGSGAGAF LASTFVFSDDL RDSVN  
N  
VILQSGSPTMQMLYQSRTSAMTAAKRLGTDLQCNTDNMTLLVSCLRQADAHVLMQSF FKKIDGFQYRMVTVMDDGIM  
KKD  
PFDYLENQDINNISVIVGTNKDEATETLMKILPRDFNRDVSQPPVFS DRKMSETLRNFLGMLYKPSYLPYATTLVNILYD  
YDKMKLTDNSSLAFLSPLALDIAYTCPMRTLIEAYTGLPNDVYVYSFNHGP DNTSLPAWVGATHRAEIPYIFRTSNRFSP  
REIALSERMQRMWANFARHGTPTPANITRTQWPVYSWL VSEQMMVFTENA  
>Unigene24585\_HaMG -1 439 3198  
MKHLGFGDQERLGIYSLVAAILHLGNVAFEDNADDAKGGCKVVKTAEDSLQTTASLLGVDKEELRDSLISRVMQATRG  
GT  
KGTAIKVPLKTAEASNARDALAKSVYSRMFDYIVDVVNKSIPFGSSVSYIGVLDIAGFEYFQVNSFEQFCINYCNEKLQQ  
FFNDRILKEEQLLYEKEGLGVKKIEWTDNQDCIELIEAKAVGIFDLLDEESKLP TPKYDHFTSEVHARNKDHFRLSVPRK  
SKLKSHRDVRDNEGFLIRHFAGAVCYHTTQFIDKNNDALHASLEFLIQDTKNPLLKLFQDTGTTAGKLN FIVSGSKFRK  
QLGLLMDKCLKATGTNFIRCIPNAKMVDHMFEGANILSQLQCAGMVSVDLDMQQGYPSRTQFSELYNMYKKYLPPELS  
RL  
DPRLFCKSLFKALGLDDKDFKFGMTKVFFRPGKFAEFDQIMKSDPESLAELVKKVKKWLLQSRWKAQWCALSVIKLK  
NK  
ILYRASKFVVLQSNIRMWRHARKFKPKFETIQVKVKA IQAQRSLLDQLVSQLKKDKEQAVKAIKELETSQENFIRQIRTTD  
LSRAEMVRMTNTLATNMDVLLNDMKKKLEEEKSREEQERLRKIQEQMEKERKKREEDERIRKEEEEEQRKKKAELERR  
RQK  
EEEEERKQEAEDKKSAAVLQKQLDEQAKENASRQAAL EQEHRDRELALRLAMEDQNQVEDITLPPPLPRSAAVVAAR  
QVA  
ASKKFDLTKWKYAE LRDAINTSCDLELLDACREEFHRR LKVYH SWKTKNKKRADKGDQDQRAPHSVEATGQQLQVAS  
TVA  
SKGTSGEQRYFRIPFARPADEYREESKHGMWYSHFDGQWIARQMELY PDKTPVLLIAGVDDLQMCESLDETGLTQK  
KGA  
EVLENEFE EEWNKHGGQALLKKHAQKVASKFLQK KLGVAK  
>CL2369.Contig2\_HaMG -2 266 1201  
MKGLGLGVVLACWVA AVELTPETTEHGAHKPSHYVDGLHNPVFDQEALLGSHKIDELAELDPEVRIKRLRN LAKSHDT  
NN  
NDVIEAHELKEWVLQSF RMLDREEAMEKLEDEDEDENG DGKVT FSEILQKQYGYTEDDLEDVQKEEEEEDEGGSNVIQLIT  
DD  
RKRFAAADVNSDGVLEEA EYIAYFQPYDFPYMYE VEMERAMKDLDKDADGYVSIQEFIGENLDEETRLSEMTNFEELD  
QD  
KDGKLTPEELRPWALPDNNDAADEEVEHLLTLCDEDKDGHLSIDEIVSKEEEFLRSTATDYGR TLHFVKDEL  
>CL1207.Contig1\_HaMG +1 10 1326  
MATTANRLLRVARLGFPSALIKRLESTTAASIAASKQKQPGSGENVEFDWRDALNLEGCLTEEEIIIRDQFRAYCQGKLM  
PRIIQANRQEYFDPTIMREMGEIGALGPTIHGYGCAGLSSVGYGLLARELERVDSSYRSAMSVQSSLVMWPISAYGSEE  
Q  
KQKYLPQLAQGNLIGCFGLTEPNHGSNPGGMETTAKY NKQRKSYVLNGTKSWITNSPIADV FVWVWAKSDVDSSRIRGF  
IL  
ERSMAGLTPKIEGKFSLRASITGQIVMEDVEVPEENLLPNVSGLSGPFGLNNARFGIAWGS LGAAEFCL ETARQYTL

D  
RIQFGRPLAKNQLIQKKLADMVTEISLGLFGCLQVGRLLKQDGLATPEMVSLKRNCGKSLDIARQARDMLGGNGISDE  
Y  
HVIRHVMNLESVNTYEGTHDIHALILGRAITGLQSFTSD  
>CL3253.Contig1\_HaMG-2 98 2671  
MSYDSDIVVRSSKVSRTYNVYRGASPSSQSRLRELEDALDSERDGRRLRAEKLVAELQFRYEQVQDSLDEQGGV  
ASQ  
QSEIRKQREQEVNKLKRDIELLNAQFEQTESSLRKRHQEAVNDLSDQLEHLQKQKIRIEKEKQTLVIELDSLQGLDSVS  
KIKASIESKLDASEQANIRLKGQVDDLARQLNDLTSKARLTQENFDLQHQQVQELDSSNAALSKAKSQLQNQNDLRRN  
L  
DDESRQRQNLQVQLAGLQSDYDSINARIEEEVESSNSLRAQLSKVNTEYSTLKARFDKEIQLKTEELEETRRKLTARIN  
E  
LEDVAEQLRVKNSGLEKTKAKLTIEIREITIELENTQIIVQDLTKRNRQLENENAAALQKRVDLGAENGLRAEKANLEQ  
EVYRLRVANAELSERNDNLQRENKQLSDQLREANQALKEANRELADLRALRAQLEAERDSLAAALKDTEALRDAENK  
LA  
NANAALQKLRQDMEARLREKDEELENIRRSSARVIEELQRTLVEVESRYKTEISRIKKKFESDIRELEGALDNANRANAE  
YKQIKSLQARVKELEGLLDEERRNADDLRSQSVSERKRLALSQELEDARSLLEAADRARKNAETELQEVNIRISELTI  
YVTSLTNDKKRLEGDIGALQSDLDLDDALNQQRAAEERAERLANENARLADELQEQDNYKNAESLRKQLEIEIREITVRL  
EAEFAQREGKRLVAKLQNRVRDLEAELEAENRHVREAVANQRKAERFYKELQTEEERKQLVELQSINDQLSIRLK  
TY  
KRQIEEAEDVANLTLNKYRKAQTLIEEAESRADNAEKNLQSVRRSRMSVTREITRVV  
>CL5457.Contig2\_HaMG+2 194 1216  
MLWWRTLLVGLVHAVPLMLVVVGEFPYQRRDHLQAKYLTTTNNENVVKDRDQAEFVLSKYGYLLCHVTRRKREASR  
YIS  
SQGNDLLPSMGVGLLEGGDACDEKEVQSAIINYQRTYNLPETGELDEETKSLMSTSRCGNTDRDKDPVQEKDKNNVDI  
SS  
NTGDNAQQKRDSTGLTEDGEDNKSSLTQNHRLWKRSEQASRSRLYQVLAGEKPVARTQEYHRRHLQEYKNKIRS  
DKRT  
VFHTYQPERRKRSVLVLRNLPHEGLSGDKGTLFNKEIIRWRLTTGISTRIPVEDQRATLDLAFRMWSEVIPLRFIE  
TSSDINHVDIEIAFGKGFIF  
>CL6000.Contig1\_HaMG-1 34 945  
MQSVRKACLGLAKASKIAFASPIAHRNAVNYVPTLFSHRKYAKEAAQPKSGVTPGRVVAVIGAVVDVQFDELPSILN  
A  
LEVQNRSPRLILEVASHLGDNIVRTIAMDGTEGLVRGEPDIDIGSPIRIPVGPCTLGRIINVIGEPIDERGPIINSEHYAA  
IHQEAPAFADMSVQQEILETGIVVDLLAPYARGGKIGLFGGAGVGKTVLIMELINNVAKAHGGYSVFAGVGERTREGN  
D  
LYHEMITTKVISLTDSDTSKVSLSVYGQMNPPGARARVALSGLTVAEFFRDQEGQDVLLFIDNIF  
>Unigene35379\_HaMG+3 90 1013  
MMAPQKRTAEAEALLVGVKSILLDIDGTTTPISTEDTLLPHVRNINIESYLKFKFDNADVKAIAALRDQAAKDKEDKV  
EGVVEIPSGDATKEDVIKAVVDNVRWQLDGSRNTELKALQEFVLRDALTSKDIKAEFGDVGQMLKMLAEEGIQLHVV  
S  
SYSVDTQKLFFANTTQGDLSDFVSGYFDATSGSKTESASYKKIATEISNEPKDILYLTHVPQEAEEAVTGGRLRSALVIRP  
GNAELTDEHLQKFACIERFDELYGDEDEDDIKRFAGDNQEQDDDEDEEGDEEDEEGDDDEGDDA  
>CL1213.Contig1\_HaMG-1 19 1086  
MASAVTWLCVLANLCLLFLHVLGRDFYKVLGVS RNVDYQIKRAYRKLAKELHPDKNKDDPEAEERFQNLGAAYEVL  
SD  
PEKRKIYDKHGEEGLKQGD MRGDPFSSFFGEGFGFGGVREQDREIPRGDVLVLDLHVTLEELYNGNFVEVLRYKPV  
AKA  
AKGKRKCNCRTEMVTQQLGPRFSMTQQQVCDECPNIRFVPEEKLEIEIEPGMRDQGEYPFVAEGEPHIDGEPGDL  
RFR  
IQQTKHARFERRGDDLYTNVTVTLADALTFEMDIKHLDGHKVHIVRDKITWPGARMRKNPNEGMPNYENNNVKGSLFIT  
F  
DIEFPKGVLTEDKKGIKEILKQENKQTVYNGLQGY  
>CL6818.Contig1\_HaMG+2 41 1360  
MLALSGLITLLCLSSVHGFYGPDDVIELTPSNFHKNVIDSKELWLVFYPWCGHCQSLTPEWKAASALKGVVKGVA  
V  
NADEHQGLGGQYGVQGFPTIKIFGADKFKPTDYQGGRTAQGIVDAALNAANS MVKARLSGKSGSSDGGDRGRSSGS  
GSGD  
SKDVIELTDSNFESLVLSQDMWLVEFFAPWCGHCKNLAPHWASAASELKGKVKLGALDATVHTV MANRYGVRGYPT  
IKF  
FPSGKKDGEAVEYDGGRTSGDIVRWALDKLAENIAPPEIYQLTDDSVLHTACDEHQLCIVSVLPHILDCQSKCRNDYLN  
V

LRKLADKYKKQMVGWIWAEAGQQLTLEESLDIGGGFYYPALAAVNSRKMKFSLFRGSFSETGINEFLRDLSYGKGSTAP  
LK  
NAKLPNVEKVEPWDGKDGVLQEDEYDVSDVELEDLDKDEL  
>Unigene32024\_HaMG +3 3 902  
MEESDTESEEDVTSDDDDSDVDDTSDVNSSGQTNEQIKDNTACVLTDFTLSCLNMDADTLSSALQTYQRVCRGLDT  
DSV  
PHRSIAAFISDDFTPWWEVFTKSNRSKPVSTLPADPNTDGDYDTSWFKNSVAGSSWLCGSPSTSVTLASETPKVED  
IS  
VPSIYDNKPAPKGRALKRERKAERDQTLGKKWFGMKAPVVDEKLNKDMELIQMRSVLDPKRFYKHRDRKGLPKYFQ  
MGT  
VVDEGSDFYSSRLTKKQRKQTLVQELMADANIRQYNKKKYAELQQKMRGKKTFRSKPKKQ  
>Unigene33735\_HaMG +1 43 273  
MYQRDTFRTVFEHFMPVTNVHVKMVRLLVKYLHCFANSTMSVVTSSYQSTNQSPLDQFKRSLIPHVLPKPLYTFLRHA

**File S2 – List of LDA precursor and buccalin-like precursors.**

>Unigene32781\_HaMG\_Cornu aspersum

MLRKTPLQQLLLLALVLTNLQISASDVSDTADSPDAFNITASKDAADIEHAVLSAREDQDSEDPDDVEKRLDSYGFY  
GIGKRRVDRFLFAEGIGKRQLDPFSFAGHLWKRINNTGGSGNRRLDKNGFAGQIGKRSLDPFGFAGQLGKRGLDKS  
GSAGAIEKRQLDTNGLAGGIGKRQANIKIDIYGFAGGIGKRSEEDGFKYDDIDDVEAESEDRHVDKRSALDKQVHVI

>Unigene4526\_Mga\_Theba\_pisana

MPRNTQLQQLLLLALVLTNLQISASDVADTADNPDAFGDNSAVNDQVNLEHSALYAPECQDSEDPDEVEKRLDGYG  
FYGGIGKRRVDRFSFTGGIGKRQLDPFSFAGHLGKRVICSTEYFDSSSKRKIDKFGFAGGIGKRRLDRFGFAGQLGKRR  
LDYGFAGGIGKRRLDYGFAGGIGKRPLDYGFAGGIGKRRLDRFGFAGGIGKRRLDRFGFAGGIGKRSEEDDFTYDD  
IDDVKAESENDSHVDKRSALDKQVGSISS

>Draft\_genome\_scaffold25752.1\_Pinctada fucata

MDPYMFRGYLGKRMDSRMFSGQLGKRALDRKMFISQLGKRDLNRMFFGRLGKRMDYRMFSGQLGKRGLDNHMFVG  
HLGKRLIPVRNSAYYYRYGSRGRPIYTQTRGMDRFSFAARLGRDAQNGTLTFSRQVFFP

>EST CU988287.1\_Crassostrea gigas

MWSTNYATTVFGFFCFVQVFLTVSQHISSHNDYTKHLENIKFLNKEAEISPKQPADDDVDFGNSDTADDLSDLTEEE  
KRALDRYSFSGSLGKRGLDRYSFYGGGLGKRALDRYGGGLGKRALDQYGFAGSLGKRALDRYSFMGGLGKRKLDQ  
YGFAGRLGKRALDRYGFVGTGKRKLDQYSFMGNLGRRLDSHRYFGSLGKRALDRYGGGLGKRADTLGNSQENI  
QGADKDEKFEQKRLYPYWYYRQGGSPIYTQTRGIDRFSFAARLGRR

>EST JZ189883.1\_Crassostrea rhizophorae

MRSYNYAKTVFGCFCLIQFLCAVSQHISSHNDLTKHLESIRFLNRQTESSPKHTDDDDTDDFRNSDISEDLSDLTDEEK  
RALDRYSFYGGGLGKRGLDRYNFYGGGLGKRALDKYGGGLGKRALDRNIFSGSLGKRALDRYSFMGGIGKRKLDYR  
FAGRLGKRELDYGFAGSLGKRRLDQFGFMGGLGKRRLDSHRFIGGLGER

>ESO97292.1\_Lottia gigantea

MAARKHELVLVLTSLVLCFVSSIVGDPNVPSDSQDNSALTQDDFAKRGMDKFGFAGGVGKRGLDKFGFTGQLGKRDM  
SFGFAGQLGKRGLDQYGFAGQLGKRGLDQYGFAGQLGKRGLDQYGFAGQLGKRGLDQYGFAGQLGKRGLDQYGF  
GQLGKRGLDQYGFAGQLGKRGLDQYGFAGQLGKRGLDQYGFAGQLGKRGLDQYGFAGQLGKRGLDQYGFAGQLG  
RGLDQYGFAGQLGKRGLDQYGFAGQLGKRGLDQYGFAGQLGKRGLDQYGFAGQLGKRGLDQYGFAGQLGKRGLD  
KYSFLGAGIGKRSVKNTAGIKKDDA

>EST ES573332.1\_Lymnaea stagnalis

MDRYGGFAGGIGKRRLDRFGFYGGIGKRDRAGGFDDWNGATEDLEKRQLDPFGFSGGIGKRKMDRFSFTGGIGKR  
RYGFSGGIGKRRLDRFSFTGGIGKRRLDRFGFAGGIGKRGLDKFGFAGGIGKRRLDRFNFAGGI  
GKRTPDDFEDNVDFEKQADEVRELDKRSTDAESHMVA

>NP\_001191649.1\_Aplysia californica

MAHHRGHRHILLYVSLALSGLALAEDATDPSDDTGSFDDVEAVSEEADLDPYSMSQELNKRPNVDPYSYLPSVKGKRA  
FDHYGFTGGGLGKRKIDHFGFVGGGLGKRQIDPLGFSGGIGKRYDSFAYSAGLGKRGMDSLAFSGGLGKRGMDSLAFSG  
GLGKRGMDSLAFSGGLGKRGMDSLAFSGGLGKRGMDSLAFSGGLGKRGMDSLAFSGGLGKRGMDSFTFAPGLGKR  
GMDSLAFAGGLGKRMDGFAFAPGLGKRMDSFAPGLGKRGMDSLAFAGGLGKRMDSFAPGLGKRMDSFAPGLGKRMDSFAP  
GLGKRGLDRYGFVGGGLGKRGMDHFAFTGGLGKRDSGEASGDLEEGKRGLDAYSFTGALGKRGLDRYGFVGGGLGKR  
GMDDFAFSPGLGKRMDSFMFGSRLGKRGMDRFSFSGHLGKRKMDQFSFGPGLGKRGFDPHYGFTGGIGKRGFDHY  
GFTGGIGKRQLDPMFLFSGRLGKRSSSEQEEEDVQRQVEKRSTTEEQSSKSL

>EST EY418574.1\_Aplysia kurodai

MIKKNKGHRQFLVYVTLALSGLALAEDATDPNHDTGSFDDVGAVSEEADLDPYSMSQELDKRPNVDPYAYLPSVKGKRA  
FDHYGFTGGIGKRKIDHFGFVGGGLGKRQIDALGFAGGIGKRYDSFAYSAGGLGKRGMDSLAFSGGLGKRGMDSLAFSG  
GLGKRGMDSLAFSGGLGKRGMDSLAFSGG

>gil54424735lgbICV548323.1\_Biomphalaria glabrata

MGIGKRDDEEDGDLEDVYEKRRIDPFAFSGGIGKRRLDRFSFAGGIGKRGIDRYGFVAGIGKRRLDRFGFSGGIGKRG  
DRFNFAGGIGKRPIDRFSFAGGIGKRGFDRYGFYGGIGKRPFDRYAFAGGIGKRPIDRFGFYGGIGKRRLDRFSFAGGI  
GKRIPDLEEVSAEELSEAANDVEKRSVPSAKSEKETVKST

>gil262323617lgbIGT274785.1\_Haliotis asinina

MDKFGFASQLGKRGMDKFGYASQLGKRGIDNFGFASQLGKRAMDKFGFASQLGKRPMDKMGFAGQLGKKAADRFA  
FASQLGKRGLDHYGFMGQLGKRGMDSFAGFAGQLGKRQVVDLSLGFSGQLGKRGMDDEYGFAGTLGKRGVDSLGFAGQ  
LGKRRMDSYGFNAQLGKRGDGDSIGKRQIDSRMFGARLGKRMDKSFAGQIDKRSA

>gil145890980lgbIES396562.1\_Mytilus californianus

MMNYYNTRKMDMYRFHGALGKRQNEQPIPEDQFEKRKIDYARFLGSLGKRATFDDSYPEVERRSKYDRFMFSPSLGK  
RYVDPQESIEADKRRLDRFSYFGNLGKRGMDKLSYFGSLGKRGLDRLSFFGGLGKRRLDRLSYFGGLGKRSGDTER  
ES

>gil238640991lgbIFL490934.1\_Mytilus galloprovincialis

MVSSVQFRCVASAVIYLILLCNELVCQESTELSKDDELLQYKRKMDMYRFHGSLGKRQDEQPIPEDQFEKRKIDYARFL  
GSLGKRANFDDSYDPVERRSKYDRYMSAPSLGKRYVDPQETMEADKRRLDRFSYFGNLGKRGMDKLSYFGSLGKR  
LDRLSFFGGLGKRRLDRLSYFGGLGKRSDRDERYNNIDSADNHNLVDVNDILPAR

>gil332167909lgbIAEE25639.1 DLamide neuropeptide precursor Platynereis dumerilii

MTHSIALQKQASKFTGKLRTNFFAQSSSYQSMAAGRQCLLTALVIFCACSQTHALNENSVPLEKEDLLLTGDDEKRYA  
FNADLGKRSQLEELYEDEDADDKRSYGFRSDLGKRMMGFNADLGKRFAAFNTDLGKRYYGFNNDLGKRYYGFNNDL  
GKRYSGFRADLGKRYMGFNADLGKRFSSFRADLGKRLRDLDESHKRYSSFRADLGKRSVPSEDLDKRFMFQRDLGK  
RYSSFRADLGKRAMFRGDLGKRFSDSSDDA

>gil13516405lgbjIBAB40308.1l minor tachykinin precursor Callistoctopus minor

MYAIRVSSRRISCFLQVFLLFSGLLNCLQWGAFANYLPKQLEPQQKNYELDSNNVNSGRYPDHEEIMEFLRKLGLLAN  
NIESRASMDTDESESAERMKKVNPYSFQGSRGKRLNANSFMGSRGKRVFANDKRTVSANAFIGSRGKRLVTLLGEKP  
QLTDLTGVIDKSDALAFVPTGRGRSQSTNDELAYMGPMLLREGRRVNSLSFGPTRGKKYSAVDFIGSRGKKSTLEDYM  
LNNQNDSPKENYERRASLHNTFMPSRGKRST

>gil49532658lgbjIBAD26598.1l tachykinin-related peptide precursor Octopus vulgaris

MYAIRVSSRRISCFLQVSLLFSGLLNLTQWGAFASYLPKQLEPQQKNYELDSNNVNSRYPGHEEIMEFLRKLGLFLAN  
TIESRASIDTDESESAERMKKVNPYSFQGTGRKRLNANSFMGSRGKRVLVNDKRTVSANAFVLSRGKRLVTLLGEKPQ

LNDLTGVIDKKSDALAFVPTRGRSQSTNDELAYTGPMLLREGRRMNSLSFGPPKGGKYSPLDFIGSRGKKSTFEDYVL  
NTQNDSPKENYERRASLHNTFIPSRGKRST

>Cern\_Brain\_C3P9AACXX\_CGATGT\_L007\_R1\_(paired)\_contig\_362] *Cernuella virgata*

MLTKTIPRQLLLLALVIACQTSASDVDPDASDSTDAIGDINEPADINPLALLAPEDQDPETAEEIEKRKIDRYGFYGGIGKR  
RVDRYSFAGGIGKRQLDPFSFAGHLGKRGIDKFGFSGGIGKRRIDKFGFTGGIGKRRLDGLELNGQLGKRRIDKFGFAG  
GIGKRRIDSYGFAGGIGKRRIDRFAGGIGKRRIDRYGFAGGIGKRRIDRYGFAGGLGKRSEDDNVVYDDVDDVEAAE  
EEADRHVGGKRSAHAETKAASVLS

>Unigene3893\_CNS\_Buccalin1\_Cochlicella\_acuta

MFTKTLPQQLLLVALILTCSHFSASDVADSTDSLSETNALNEQSELGPLALFTPEDRDSEATEDVDKRRKIDRYGFYGGIG  
KRRVDRFGFTGGIGKRELDPFSAFAGHLGKRGIDKFGFYGGIGKRRLDKFGFTGGIGKRRLDKFGFTGGIGKRRIDKFGF  
AGGIGKRRIDSYGFAGGIGKRRIDRFAGGLGKRRIDRYGFAGGIGKRSDDDNIMYDDTDDVAVAEAEERHVDKRSA  
QVGHQAAAVSS

## **A "Love" Dart Allohormone Identified in the Mucous Glands of Hermaphroditic Land Snails**

Michael J. Stewart, Tianfang Wang, Joris M. Koene, Kenneth B. Storey and Scott F. Cummins

*J. Biol. Chem.* 2016, 291:7938-7950.

doi: 10.1074/jbc.M115.704395 originally published online January 27, 2016

---

Access the most updated version of this article at doi: [10.1074/jbc.M115.704395](https://doi.org/10.1074/jbc.M115.704395)

### Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

### Supplemental material:

<http://www.jbc.org/content/suppl/2016/01/27/M115.704395.DC1.html>

This article cites 38 references, 12 of which can be accessed free at <http://www.jbc.org/content/291/15/7938.full.html#ref-list-1>