



ConoServer: updated content, knowledge, and discovery tools in the conopeptide database

Quentin Kaas, Riley Yu, Ai-Hua Jin, Sébastien Dutertre, David J. Craik

► To cite this version:

Quentin Kaas, Riley Yu, Ai-Hua Jin, Sébastien Dutertre, David J. Craik. ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Research*, 2011, 40 (D1), pp.D325-D330. 10.1093/nar/gkr886 . hal-02306909

HAL Id: hal-02306909

<https://hal.science/hal-02306909>

Submitted on 7 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ConoServer: updated content, knowledge, and discovery tools in the conopeptide database

Quentin Kaas, Rilei Yu, Ai-Hua Jin, Sébastien Dutertre and David J. Craik*

Division of Chemistry and Structural Biology, Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia

Received September 13, 2011; Accepted October 2, 2011

ABSTRACT

ConoServer (<http://www.conoserver.org>) is a database specializing in the sequences and structures of conopeptides, which are toxins expressed by marine cone snails. Cone snails are carnivorous gastropods, which hunt their prey using a cocktail of toxins that potentially subvert nervous system function. The ability of these toxins to specifically target receptors, channels and transporters of the nervous system has attracted considerable interest for their use in physiological research and as drug leads. Since the founding publication on ConoServer in 2008, the number of entries in the database has nearly doubled, the interface has been redesigned and new annotations have been added, including a more detailed description of cone snail species, biological activity measurements and information regarding the identification of each sequence. Automatically updated statistics on classification schemes, three-dimensional structures, conopeptide-bearing species and endoplasmic reticulum signal sequence conservation trends, provide a convenient overview of current knowledge on conopeptides. Transcriptomics and proteomics have begun generating massive numbers of new conopeptide sequences, and two dedicated tools have been recently implemented in ConoServer to standardize the analysis of conopeptide precursor sequences and to help in the identification by mass spectrometry of toxins whose sequences were predicted at the nucleic acid level.

INTRODUCTION

Peptide toxins expressed by cone snails, or conopeptides, display a high level of chemical diversity, allowing them to potentially target receptors, ion channels and transporters of the nervous systems (1–3). Conopeptides, and especially

their disulfide-rich subclass referred to as conotoxins, attract considerable interest in both fundamental research and applied sciences, as evidenced by approximately 4000 articles published (based on a search in NCBI PubMed using the keyword ‘conotoxins’). Because of their exquisite specificity for receptor subtypes, conopeptides are valuable tools in neurological studies (4–6) and several are being developed as drugs or drug leads (7–10). The most advanced of these, MVIIA or ziconotide, is a Food and Drug Administration (FDA)-approved analgesic (11) and is more potent than morphine without developing tolerance. Two other conopeptides have entered human clinical trials for the treatment of neuropathic pain (12) and others are in preclinical evaluation. Furthermore, numerous fundamental biological studies focused on understanding the maturation of the venom (13), the influence of environment or cone snail development stage on conopeptide expression (14–16) and the phylogenetic relationships between toxins (17,18) have been published.

ConoServer (<http://www.conoserver.org>) is a database that aims to organize information on conopeptides for easy and convenient access to conopeptide discovery, structure and activity data as well as data on venom evolution. Interest in conopeptide sequence and structural data prompted the creation of ConoServer in 2007 (19), and it has been a very popular website, with an average of 300 hits per day currently recorded. ConoServer has been recognized as a valuable source of annotations by UniProt, and since January 2010 links are formally exchanged between UniProt-KB and ConoServer. Two important missions of ConoServer are to help organize knowledge on conopeptides and to provide tools to help in the analysis and comparison of conopeptides.

Conopeptides have been categorized in the literature using several classification schemes; the three classifications used in ConoServer are: the gene superfamilies classification that is based on similarities in conopeptide precursor sequences, the cysteine frameworks classification based on patterns of cysteines in the mature peptide domain and the pharmacological families classification that categorizes conopeptides according to their activity.

*To whom correspondence should be addressed. Tel: +61 7 3346 2019; Fax: +61 7 3346 2101; Email: d.craik@imb.uq.edu.au

ConoServer helps to keep track of the use and evolution of these classification schemes, and this function recently facilitated the identification of inequalities in data collected among different clades of cone snails and helped to define new directions of research (20). ConoServer has also proven to be a valuable tool to avoid the unintended reuse of names, since currently accepted nomenclature for conopeptides requires knowledge of the order of peptide discovery. As well as continuing to provide useful classification and nomenclature functions, ConoServer now provides tools to analyze newly discovered conopeptide precursor sequences and deal with increasingly complex venom mass spectrometry data on mature peptides.

This article describes significant updates to ConoServer that have been implemented since its initial publication in 2008 (19). General statistics on database content and descriptions of new types of annotations introduced into ConoServer are presented first and a new interface implemented to improve accessibility to ConoServer content is then described. Finally, two bioinformatics tools that help process transcriptomics and proteomics data are described. The analysis of massively parallel sequencing and mass spectrometry data is challenging, and the new ConoPrec and ConoMass tools help to meet this challenge by addressing specific issues related to conopeptides.

DATABASE CONTENT

ConoServer annotations associated with individual conopeptide sequences are entered semi-automatically and manually. An annotation system performs most of the repetitive tasks, but the resulting outputs in all cases are subjected to manual reviewing before being approved and published. The majority of the sequence and three-dimensional structure data are retrieved from publicly available databases, including GenBank (21), UniProt-KB (22), the Protein Data Bank (23) and the Biological Magnetic Resonance Bank (24). Manual curation of the peer-reviewed literature provides additional entries, which are therefore unique to ConoServer. Conopeptides are expressed as prepropeptides (25), and their corresponding mature peptide is predicted using ConoPrec for cases where it was not identified in the literature. As of September 2011, ConoServer provides information on 1180 mature conopeptides. However, with more than 500 species of cone snails (26) and estimates of 200–1000 unique conopeptides per species (27), the number of known peptides cataloged in ConoServer is only a small fraction of the potential pool of wild-type conopeptides. ConoServer will need to be regularly updated and improved to cope with the increasing number of sequences.

ConoServer now provides sequence/structure/activity relationships information that is of particular interest for drug design studies. Examples of bioactivity data that are now provided include measures of IC_{50} , K_i , K_d and percentage of inhibition of ion currents in various electrophysiological assays. Besides native conopeptide sequences, ConoServer contains information on 338 synthetic variants, which have been chemically synthesized

to study the receptor specificity and stability of conopeptides with potentially interesting pharmaceutical properties. ConoServer catalogs 95 three-dimensional structures of wild-type conopeptides and 42 structures of synthetic variants. The majority of these structures have been determined by nuclear magnetic resonance (28). Finally, ConoServer describes 1288 patented protein and 737 patented nucleic acid sequences.

New types of annotations related to the discovery and evolution of conopeptides are now available in ConoServer, including a more extensive description of organisms, information on how mature peptide sequences were identified and the analysis of precursor sequences. The geographic location and the diet (mollusk, worm or fish) of specific cone snails are new features that are retrieved from the Conus Biodiversity website (<http://biology.burke.washington.edu/conus/>) or from the peer-reviewed literature. Mature conopeptides are typically either isolated directly from the venom or predicted from a nucleic acid precursor. Information on the method of identification, now included in ConoServer, allows users to make a rapid assessment of the confidence of conopeptide sequences and the presence of post-translational modifications. Conopeptides are classified into gene superfamilies according to the similarity of the endoplasmic reticulum (ER) signal sequence in their precursor. For cases where the ER signal sequence is not identified in the literature, ConoServer predicts it using the new tool ConoPrec (described below). The sequences of 1120 precursors are currently in ConoServer and 16 gene superfamilies are described. In addition, 13 other temporary gene superfamily were recently introduced in ConoServer to describe newly discovered conopeptide precursors expressed by cone snails from the ‘early divergent’ clade (15,17).

ConoServer now computes statistics on known conopeptides. The statistical tables are kept up-to-date with the database content, and provide information on relationships between classification schemes, sequence conservation of signal sequence regions that define gene superfamilies, the number of conopeptides for each species and details on three-dimensional structures. As an example of the use of this information, these statistics were valuable in a recent discussion of the relationships between the various conopeptide classification schemes (20). The statistical tables also provide a convenient access link to the database content. For example, there are 18 conopeptides that are antagonists of sodium channels (μ pharmacological family), and some of them belong to the M gene superfamily. Clicking on the ‘M’ in the corresponding table gives access to the list of the 11 μ -conopeptides belonging to the M superfamily.

IMPROVED ACCESS TO THE DATABASE

Figure 1 shows the new interface that was designed to improve the ergonomics of the website. A search bar located in the page header allows users to retrieve conopeptides by name or identifier in the protein, nucleic acid or three-dimensional structure entries. A menu located below the

search bar gives access to advanced searches, web-based tools, statistics pages or to descriptions of the classification schemes. The search by references was modified to display the list of all references used in ConoServer, sorted by year and first author name. Links displayed next to each reference lead to the corresponding list of peptides and nucleic acids. The advanced search for peptides, nucleic acids and three-dimensional structures allows users to select multiple search criteria, to use sequence information and to select the fields to be displayed in the resulting list of entries. Utilization of the various conopeptide classification schemes requires knowledge of their definition, and easy access to tables defining the different

classes is now directly provided from the top menu. The classification scheme tables are provided with textual explanations that clarify the definitions in use in ConoServer. The result lists can also be used to align sequences using CLUSTALW (29), draw LOGO representations (30) or generate phylogenetic trees using PHYLIP (31).

The complete set of information on each conopeptide sequence or structure is displayed on ConoServer cards. A partial view of the card for conopeptide MrIA is shown in Figure 1. New features displayed on the cards include cone snail geographic locations and diet, biological assay data and a list of synthetic variants. A photograph of the shell of the corresponding cone snail is also shown when

ConoServer - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.conoserver.org/

Search a Protein name ☐ exact match Search Clear

HOME ABOUT SEARCH TOOLS STATISTICS CLASSIFICATION

MrIA (P01384) Protein

General Information

Name	MrIA
Alternative name(s)	CMrVIB,Mr10.1,Xen2174,mr10a
Organism	Conus marmoreus
Organism region	Indo-Pacific
Organism diet	molluscivorous
Protein Type	Wild type
Protein precursor	MrIA precursor (1739)
Notes	Targets the norepinephrine transporter.

Classification

Conopeptide class	conotoxin
Gene superfamily	T superfamily
Cysteine framework	X [connectivity I-IV, II-III]
Pharmacological family	chi conotoxin

Sequence

NGVCCGYKLCHOC

Modified residues

position	symbol	name
12	O	4-Hydroxyproline

Sequence evidence protein level

Average Mass 1408.64

Monoisotopic Mass 1407.52

Isoelectric Point 10.45

Extinction Coefficient [280nm] 1490.00

Figure 1. ConoServer interface and protein card for conopeptide MrIA from *Conus marmoreus*, shown as a representative example of the updated interface. The top of the website displays a search bar and a menu that allows users to navigate between textual information pages, search forms, statistics tables and descriptions of the various conopeptide classification schemes.

available. These images are either in the public domain or provided with permission by collectors, as indicated on the website. Activity data include the source organism of the receptor subtype tested, the agonist and its concentration, the competitive inhibitor and its concentration, the Hill coefficient, notes and a peer-reviewed or a patent reference. Sequences of synthetic variants of each entry are also listed. For protein precursors, the signal sequence region and mature peptide region are now highlighted in the sequence.

ConoServer data are available for download in XML format using a link located in the 'Tools' menu. Protein, nucleic acid and structural data are described in separate files, whose contents are synchronized with the database.

CONOPREC: ANALYSIS OF CONOPEPTIDE PRECURSORS

ConoPrec provides a standardized analysis of conopeptide precursor sequences. This tool is used internally to analyze some of the data in ConoServer and is available to users via the website. Modern transcriptomic techniques produce a deluge of contig sequences that need to be analyzed. The high sequence variability of conopeptides renders classical analysis by sequence alignment inefficient for contig identification, and ConoPrec was designed to help select and analyze contigs coding for conopeptide precursors. This web-based tool has been already employed in a recent publication analyzing transcript sequences from *Conus californicus* (15).

Users can submit to ConoPrec a single nucleic acid or protein sequence, or, alternatively, can upload a file containing a set of sequences in FASTA format. The submission of a single sequence produces a detailed on-line output, whereas batch submission of sequences produces several output files, in XLS, Comma Separated Values (CSV) and text format. The outputs for each submitted precursor include the identification of sequence regions, classification according to the three classification schemes, identification of the most similar sequences in ConoServer and predictions of potential post-translational modifications of the mature conopeptide.

In the case when a nucleic acid sequence is submitted, ConoPrec identifies the most probable open reading frame (ORF) on the basis of the presence of a leading methionine, ORF length and Kozak consensus sequence statistics. The ORF is then translated and the precursor protein sequence is further analyzed. The ER signal sequence region is identified using the signalP algorithm (32). The gene superfamily is determined by sequence similarity to ER signal sequences from conopeptides already annotated in ConoServer. If no signal sequence has an identity >90%, the gene superfamily is not assigned. In that case, and if a single sequence was submitted, the maximum percentage of identity within each superfamily is provided to the user. The boundaries of the mature peptide region are then determined using sequence patterns that predict the cleavage sites of endopeptidases (typically proprotein convertases that are widely implicated in protein

processing) (33) and two exopeptidases that are known to be involved in mollusk protein maturation (34,35): carboxypeptidase E, which cleaves C-terminal lysines and arginines, and peptidylglycine α -amidating monooxygenase, which cleaves a C-terminal glycine. Three types of post-translational modifications can be predicted: C-terminal amidation, pyroglutamylation and γ -carboxylation. Since α -amidating monooxygenase performs C-terminal amidation, this modification is predicted when activity of this exopeptidase is predicted. The modification of an N-terminal glutamine or glutamate into pyroglutamic acid occurs spontaneously, and can therefore be accurately predicted. The modification of glutamate into γ -carboxylic glutamic acids is predicted using a recognition sequence pattern matching part of the proprotein sequence (36). Specifically, the study of Czerwec *et al.* (36) was extended here using current ConoServer data, leading to a refined pattern used in ConoPrec: [KR].[2,3][ACGILMFVS].[3,4][KRN], where alternative amino acids at a given position are between square brackets, a dot followed by curly brackets denotes repeats of any type of amino acid, and the possible lengths of these repeats are indicated between the curly brackets.

CONOMASS: ANALYSIS OF PROTEOMIC RESULTS

The ConoMass tool was implemented in ConoServer to match peptide masses predicted from transcripts with a list of masses obtained experimentally by proteomics analysis of cone snail venoms. The high frequency and variability of post-translational modifications in conopeptides is a major challenge for the success of this task. Indeed, besides disulfide bond formation, 13 other post-translational modifications have been so far identified in wild-type conopeptides (20). ConoMass analysis is divided into two steps: (i) computation of the masses corresponding to all possible modifications of conopeptide sequences predicted from transcripts; and (ii) identification of these predicted masses in experimental mass spectra. This stepwise approach allows users to compare several sets of mass spectrometry data to the same list of masses derived from transcript sequences, whose generation is the most time consuming for a large-scale contig database.

Users can upload a list of mature conopeptide sequences in FASTA format to ConoMass or submit a single sequence using a similar interface to that of ConoPrec. All or some of 12 post-translational modifications can be selected. Among the 13 known wild-type modifications, only glycosylation is not dealt within ConoMass because it potentially generates an enormous number of possibilities. Only one conopeptide with a glycosylated serine and three conopeptides with a glycosylated threonine have been discovered so far and thus the exclusion of glycosylation from ConoMass analysis is not expected to be a significant limitation. Chemical modifications commonly employed before mass spectrometry analyses, such as reduction and alkylation of cysteines, can also be considered. ConoMass output files are available in XLS, CSV and text

file formats. The files contain the list of predicted monoisotopic and average masses, their corresponding sequences and the number and nature of the post-translational modifications. These files are kept for 2 days on the server and a session number allows users to retrieve them at any time. The session number can also be used in the second part of the tool without having to download or upload result files.

On the mass spectrometry comparison page of ConoMass, users are requested to provide mass spectrometry data files, which can be in CSV, Rich Text File (RTF) or text formats. After uploading the files, users must select a column from the mass spectrometry files, identify the masses as monoisotopic or average and select an adequate mass correction. In the second part of the page, the list of predicted peptide masses can be uploaded using the text format generated in the first step of ConoMass. Alternatively, the session number of the first step of ConoMass can be provided to indicate a list of predicted masses still stored on the server. A precision indicative of a mass match should also be provided. The computations of the ConoMass tool are submitted to a queuing system that prevents overload of the server capacity. The output of ConoMass is a list of post-translationally modified conopeptide sequences whose masses were identified in a list of experimental masses derived from crude venom mass spectrometry analysis. Three other tools available in the 'Tools' section of ConoServer help to clean and bin the proteomic data and compare results of different mass spectrometry experiments. These three tools are provided as a complement to ConoMass but are not conopeptide specific.

ConoMass should significantly facilitate the validation of integrated venom strategies for the accelerated discovery of novel conopeptides. However, the limitations of this tool will be directly linked to the accuracy of the mass spectrometry data for efficient matching results as well as to the ability of the users to provide correctly predicted mature peptide sequences.

CONCLUSIONS

The current version of ConoServer provides users with a new interface, new tools for conopeptide analysis and significantly enhanced information content in the rapidly evolving field of conopeptides. It is hoped that these and other ongoing improvements will further enhance the use of the database and, in doing so, facilitate conopeptide research.

FUNDING

Australian National Health and Medical Research Council NHMRC (Grant no. 631457); Australian Research Council (DP1093115). Funding for open access charge: University of Queensland.

Conflict of interest statement. None declared.

REFERENCES

1. Terlau, H. and Olivera, B.M. (2004) Conus venoms: a rich source of novel ion channel-targeted peptides. *Physiol. Rev.*, **84**, 41–68.
2. Janes, R.W. (2005) alpha-Conotoxins as selective probes for nicotinic acetylcholine receptor subclasses. *Curr. Opin. Pharmacol.*, **5**, 280–292.
3. Olivera, B.M., Quik, M., Vincler, M. and McIntosh, J.M. (2008) Subtype-selective conopeptides targeted to nicotinic receptors: concerted discovery and biomedical applications. *Channels*, **2**, 143–152.
4. Olivera, B.M. and Cruz, L.J. (2001) Conotoxins, in retrospect. *Toxicon*, **39**, 7–14.
5. Dutton, J.L. and Craik, D.J. (2001) alpha-Conotoxins: nicotinic acetylcholine receptor antagonists as pharmacological tools and potential drug leads. *Curr. Med. Chem.*, **8**, 327–344.
6. Lewis, R.J. (2009) Conotoxins: molecular and therapeutic targets. *Prog. Mol. Subcell. Biol.*, **46**, 45–65.
7. Craik, D.J. and Adams, D.J. (2007) Chemical modification of conotoxins to improve stability and activity. *ACS Chem. Biol.*, **2**, 457–468.
8. Vincler, M. and McIntosh, J.M. (2007) Targeting the alpha9alpha10 nicotinic acetylcholine receptor to treat severe pain. *Expert Opin. Ther. Targets*, **11**, 891–897.
9. Twede, V.D., Miljanich, G., Olivera, B.M. and Bulaj, G. (2009) Neuroprotective and cardioprotective conopeptides: an emerging class of drug leads. *Curr. Opin. Drug Discov. Devel.*, **12**, 231–239.
10. Clark, R.J., Jensen, J., Nevin, S.T., Callaghan, B.P., Adams, D.J. and Craik, D.J. (2010) The engineering of an orally active conotoxin for the treatment of neuropathic pain. *Angew. Chem. Int. Ed.*, **49**, 6545–6548.
11. Miljanich, G.P. (2004) Ziconotide: neuronal calcium channel blocker for treating severe chronic pain. *Curr. Med. Chem.*, **11**, 3029–3040.
12. Halai, R. and Craik, D.J. (2009) Conotoxins: natural product drug leads. *Nat. Prod. Rep.*, **26**, 526–536.
13. Safavi-Hemami, H., Siero, W.A., Gorasia, D.G., Young, N.D., Macmillan, D., Williamson, N.A. and Purcell, A.W. (2011) Specialisation of the venom gland proteome in predatory cone snails reveals functional diversification of the conotoxin biosynthetic pathway. *J. Proteome Res.*, **10**, 3904–3919.
14. Duda, T.F. Jr, Chang, D., Lewis, B.D. and Lee, T. (2009) Geographic variation in venom allelic composition and diets of the widespread predatory marine gastropod *Conus ebraeus*. *PLoS One*, **4**, e6245.
15. Elliger, C.A., Richmond, T.A., Lebaric, Z.N., Pierce, N.T., Sweedler, J.V. and Gilly, W.F. (2011) Diversity of conotoxin types from *Conus californicus* reflects a diversity of prey types and a novel evolutionary history. *Toxicon*, **57**, 311–322.
16. Safavi-Hemami, H., Siero, W.A., Kuang, Z., Williamson, N.A., Karas, J.A., Page, L.R., Macmillan, D., Callaghan, B., Kompella, S.N., Adams, D.J. et al. (2011) Embryonic toxin expression in the cone snail *Conus victoriae*: primed to kill or divergent function? *J. Biol. Chem.*, **286**, 22546–22557.
17. Biggs, J.S., Watkins, M., Puillandre, N., Ownby, J.-P., Lopez-Vera, E., Christensen, S., Moreno, K.J., Bernaldez, J., Licea-Navarro, A., Corneli, P.S. et al. (2010) Evolution of *Conus* peptide toxins: analysis of *Conus californicus* Reeve, 1844. *Mol. Phylogenet. Evol.*, **56**, 1–12.
18. Jimenez, E.C. and Olivera, B.M. (2010) Divergent M- and O-superfamily peptides from venom of fish-hunting *Conus parius*. *Peptides*, **31**, 1678–1683.
19. Kaas, Q., Westermann, J.-C., Halai, R., Wang, C.K.L. and Craik, D.J. (2008) ConoServer, a database for conopeptide sequences and structures. *Bioinformatics*, **24**, 445–446.
20. Kaas, Q., Westermann, J.-C. and Craik, D.J. (2010) Conopeptide characterization and classifications: an analysis using ConoServer. *Toxicon*, **55**, 1491–1509.
21. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Sayers, E.W. (2011) GenBank. *Nucleic Acids Res.*, **39**, D32–37.
22. UniProt Consortium. (2011) Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.*, **39**, D214–219.

23. Rose,P.W., Beran,B., Bi,C., Bluhm,W.F., Dimitropoulos,D., Goodsell,D.S., Prlic,A., Quesada,M., Quinn,G.B., Westbrook,J.D. *et al.* (2010) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res.*, **39**, D392–D401.
24. Ulrich,E.L., Akutsu,H., Doreleijers,J.F., Harano,Y., Ioannidis,Y.E., Lin,J., Livny,M., Mading,S., Maziuk,D., Miller,Z. *et al.* (2008) BioMagResBank. *Nucleic Acids Res.*, **36**, D402–408.
25. Woodward,S.R., Cruz,L.J., Olivera,B.M. and Hillyard,D.R. (1990) Constant and hypervariable regions in conotoxin propeptides. *EMBO J.*, **9**, 1015–1020.
26. Duda,T.F. Jr, Kohn,A.J. and Matheny,A.M. (2009) Cryptic species differentiated in *Conus ebraeus*, a widespread tropical marine gastropod. *Biol. Bull.*, **217**, 292–305.
27. Davis,J., Jones,A. and Lewis,R.J. (2009) Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. *Peptides*, **30**, 1222–1227.
28. Nicke,A., Loughnan,M.L., Millard,E.L., Alewood,P.F., Adams,D.J., Daly,N.L., Craik,D.J. and Lewis,R.J. (2003) Isolation, structure, and activity of GID, a novel alpha 4/7-conotoxin with an extended N-terminal sequence. *J. Biol. Chem.*, **278**, 3137–3144.
29. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
30. Schneider,T.D. and Stephens,R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.
31. Felsenstein,J. (1989) PHYLIP – Phylogeny Inference Package (Version 3.2). *Cladistics*, **5**, 164–166.
32. Bendtsen,J.D., Nielsen,H., von Heijne,G. and Brunak,S. (2004) Improved prediction of signal peptides: signalP 3.0. *J. Mol. Biol.*, **340**, 783–795.
33. Duckert,P., Brunak,S. and Blom,N. (2004) Prediction of proprotein convertase cleavage sites. *Protein Eng. Des. Sel.*, **17**, 107–112.
34. Fan,X. and Nagle,G.T. (1996) Molecular cloning of *Aplysia* neuronal cDNAs that encode carboxypeptidases related to mammalian prohormone processing enzymes. *Dev. Cell Biol.*, **15**, 937–945.
35. Fan,X., Spijker,S., Akalal,D.B. and Nagle,G.T. (2000) Neuropeptide amidation: cloning of a bifunctional alpha-amidating enzyme from *Aplysia*. *Brain Res. Mol. Brain Res.*, **82**, 25–34.
36. Czerwiec,E., Kalume,D.E., Roepstorff,P., Hambe,B., Furie,B., Furie,B.C. and Stenflo,J. (2006) Novel gamma-carboxyglutamic acid-containing peptides from the venom of *Conus textile*. *FEBS J.*, **273**, 2779–2788.