BIOINFORMATICS APPLICATIONS NOTE Vol. 24 no. 3 2008, pages 445–446
doi:10.1093/bioinformatics/btm596

Databases and ontologies

ConoServer, a database for conopeptide sequences and structures

Quentin Kaas, Jan-C. Westermann, Reena Halai, Conan K. L. Wang and David J. Craik*

Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, 4072, Australia

Received on October 15, 2007; revised on November 26, 2007; accepted on November 27, 2007

Advance Access publication December 6, 2007

Associate Editor: Alex Bateman

ABSTRACT

Summary: ConoServer is a new database dedicated to conopeptides, a large family of peptides found in the venom of marine snails of the genus *Conus*. These peptides have an exceptional diversity of sequences and chemical modifications and their ability to block ion channels makes them important as drug leads and tools for physiological studies. ConoServer uses standardized names and a genetic and structural classification scheme to present data retrieved from SwissProt, GenBank, the Protein DataBank and the literature. The ConoServer web site incorporates specialized features like the graphic display of post-translational modifications that are extensively present in conopeptides. Currently, ConoServer manages 1214 nucleic sequences (from 54 *Conus* species), 2258 proteic sequences (from 66 *Conus* species) and 99 3D structures.

Availability: http://research1t.imb.uq.edu.au/conoserver/

Contact: d.craik@imb.uq.edu.au

1 INTRODUCTION

Predatory marine cone snails of the *Conus* genus produce a concoction of venom peptides, referred to as conopeptides, which they use to capture prey (Gray *et al*., 1988; Terlau and Olivera, 2004). With >700 *Conus* species, each with their own repertoire of peptide toxins, conopeptides form a huge library of bioactive peptides (Olivera, 1997; Olivera and Cruz, 2001). They are typically 10–40 amino acids long and contain up to five disulfide bonds (Craik *et al*., 2007). Due to their high specificity for ion channel isoforms and their high potency, conopeptides are of great interest as neuropharmacological tools and as drug leads (Adams *et al*., 1999). Prialt^{TM}, a synthetic version of a conopeptide from *Conus magus*, is an approved drug for the treatment of chronic pain. Several other conopeptides are in clinical or pre-clinical trials.

Conopeptides are synthesized as prepro-peptides, which are proteolytically cleaved to yield the mature peptide. The signal sequence is well conserved, but the mature toxin sequence, at the C-terminus of the prepro-peptide, is highly divergent. The mature peptides have a high frequency of post-translational modifications. Conopeptides are classified into disulfide-rich and -poor, into superfamilies (signal sequence similarity), into cysteine framework categories and into pharmacological families, as explained in Figure 1. With interest in conotoxin growing, a database is needed to systematize the increasing number of discovered sequences and structures.

2 DATA RETRIEVAL AND ANNOTATIONS

The sequences and structures of conopeptides were extracted from public databases, GenBank (Benson *et al*., 2007), SwissProt (Boeckmann *et al*., 2003) and the Protein DataBank (Berman *et al*., 2000), and by an extensive survey of the literature. The sequences of mature peptides were also extracted from prepro-peptide sequences.

Manual curation of each entry contributes to a high level of standardization that is necessary for efficient searches and comparisons. Wherever applicable a standard naming scheme (Gray *et al*., 1988) was used: one or two letters indicating the *Conus* species, a Roman numeral indicating the disulfide framework category and an upper case letter denoting the order of discovery. A list of alternative names found in the literature was also built. The conotoxin superfamilies were assigned based on the analysis of the signal sequence.

ConoServer currently manages 1214 nucleic sequences (from 54 *Conus* species), 2258 proteic sequences (from 66 species) and 99 3D structures. The proteic sequences are split into 450 mature peptides, 615 prepro-peptides, 34 synthetic peptides and 1159 sequences from patents. The 427 mature conotoxins are split into superfamilies as follows: 133 O, 104 A, 58 M, 51 T, 31 I, 7 L, 6 P, 6 J, 6 P, 3 D, 2 S and 1 G superfamily peptides. The superfamilies of the remaining 19 conotoxins have not yet been published.

3 INTERFACE AND VISUALIZATION

3.1 Search

ConoServer allows searches of nucleic acids, proteins and 3D structures of conopeptides based on their name, patent ID, sub-sequence, FASTA alignment, mass range, peptide mass fragments (fingerprints), classification (Fig. 1), type (mature peptide, prepro-peptide, synthetic peptide or patent) and species. The name search simultaneously uses standard names and related names (historical names, non-standard names, trade names).

3.2 Results

The results are displayed as a table whose column fields are customizable: standard names, target families, superfamilies,

*To whom correspondence should be addressed.

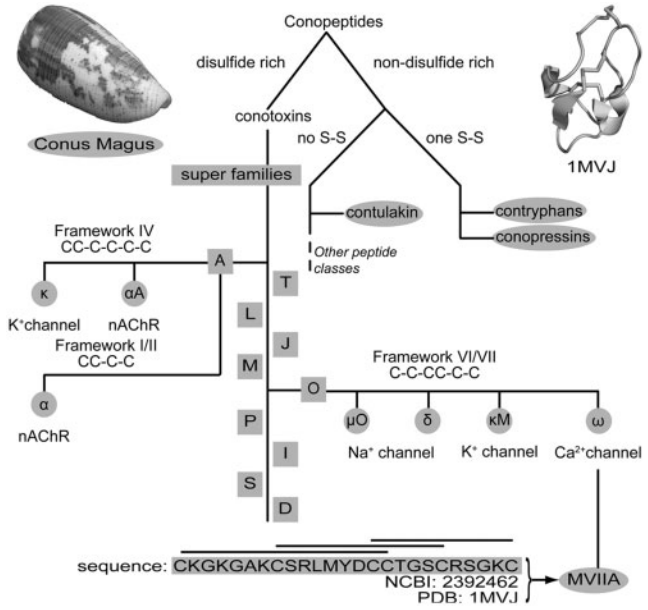

Fig. 1. Illustration of the conopeptide classification scheme for the example conotoxin MVIIA. Conopeptides are split into disulfide-rich, or conotoxins and non-disulfide-rich. Conotoxins are further divided into superfamilies (noted A to S and defined by signal sequence similarity), according to their cysteine framework pattern (noted with a Roman numeral) and by their receptor specificity (noted with a Greek letter). Two superfamily branches are expanded. MVIIA (ziconotide, Prialt), from *Conus magus* (top left) belongs to the O superfamily. Searchable terms have a grey background.

protein types, species, sequences, masses, curation notes, literature references and external links. A list of more than 400 references (linked to the corresponding PubMed abstract) is stored in the database and can be displayed.

3.3 Cards

Each element of the result list is linked to an entry card presenting four parts: (i) general information with name, classification, sequence, mass, isoelectric point and extinction coefficient (ii) literature references with links to PubMed, (iii) cross references between ConoServer cards and with external databases and (iv) tools to predict the digestion of proteic sequences and to visualize the 3D structures with the Jmol applet (http://www.jmol.org/). The visualization of sequences in cards and lists highlights the precursors in nucleic acids sequences, the signal sequence and the mature peptide in prepro-sequences, and the cysteine framework and the post-translational modifications in mature peptide sequences.

3.4 Sequence comparison

ConoServer allows comparison of sequences of entries selected from a result list. The sequences can be aligned with CLUSTALW (Thompson *et al.*, 1994) and the alignment analysed with an amino acid based colour scheme or with a LOGO representation (Schneider and Stephens, 1990) or with a distance tree computed with *protdist* and *dnadist* from the PHYLIP package (Felsenstein, 1989).

4 IMPLEMENTATION

ConoServer uses MySQL (http://www.mysql.com) and its web interface is implemented in PHP (http://www.php.net). A single XML file for proteins, nucleic acids and structures allows a common definition of the search, list and card pages. A web-based annotation interface allows efficient entry or change of data. ConoServer is freely available at http://research1t.imb.uq.edu.au/conoserver/.

5 CONCLUSIONS

ConoServer is a new database that provides standardized annotations of conopeptides. The web interface allows searching of the database using a combination of criteria that include the conopeptide sequence, classification and names. The unique display features and cross links between nucleic acid, proteic and structural data and the high quality of the annotations will hopefully make ConoServer a useful resource for researchers working on conopeptides and more broadly on bioactive peptides.

ACKNOWLEDGEMENT

Work in our laboratory on conotoxins is supported by the Australian Research Council.

Conflict of Interest: none declared.

REFERENCES

Adams,D.J. *et al.* (1999) Conotoxins and their potential pharmaceutical applications. *Drug Dev. Res.*, **46**, 219–234.

Benson,D.A. *et al.* (2007) GenBank. *Nucleic Acids Res.*, **35**, D21–D25.

Berman,H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

Boeckmann,B. *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **31**, 365–370.

Craik,D.J. and Adams,D.J. (2007) Chemical modification of conotoxins to improve stability and activity. *ACS Chem. Biol.*, **2**, 457–468.

Felsenstein,J. (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics*, **5**, 164–166.

Gray,W.R. *et al.* (1988) Peptide toxins from venomous Conus snails. *Annu. Rev. Biochem.*, **57**, 665–700.

Olivera,B.M. (1997) E.E. Just Lecture, 1996. Conus venom peptides, receptor and ion channel targets, and drug design: 50 million years of neuropharmacology. *Mol. Biol. Cell*, **8**, 2101–2109.

Olivera,B.M. and Cruz,L.J. (2001) Conotoxins, in retrospect. *Toxicon*, **39**, 7–14.

Schneider,T.D. and Stephens,R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.

Terlau,H. and Olivera,B.M. (2004) Conus venoms: a rich source of novel ion channel-targeted peptides. *Physiol. Rev.*, **84**, 41–68.

Thompson,J.D. *et al.* (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.