Protein toxins: overview of databases and other resources

Florence Jungo, UniProtKB toxin annotation program
SIB Swiss Institute of Bioinformatics, Swiss-Prot group

HESI Protein Toxins Workshop (virtual) / October 2020

www.sib.swiss
Protein toxin databases

Introduction
UniProtKB/Swiss-Prot
ConoServer
ArachnoServer
Kalium

Other resources on protein toxins
Toxin prediction tools
VenomZone
Knottin
Sources of protein toxins

**Animals**
- Northern short-tailed shrew
- Platypus
- Gila Monster
- Stonefish
- Sting ray
- Scolopendra

**Plants**
- Ricin

**Fungi**
- Alpha-amanitin

**Bacteria**
- Viruses
Specific databases for protein toxins

- ConoServer (cone snails)
- ArachnoServer (spiders)
- Kalium (toxins that target K+ channels)

Animals

Plants

Fungi
- Norine toxic and non-toxic nonribosomal peptides
  ([https://bioinfo.cristal.univ-lille.fr/norine/](https://bioinfo.cristal.univ-lille.fr/norine/))

Bacteria
- Norine

Viruses
Databases for protein toxins

Animals
- ConoServer (toxins from cone snails)
- ArachnoServer (toxins from spiders)
- Kalium (toxins that target K+ channels)
- UniProtKB/Swiss-Prot (Toxin project)

Plants
- UniProtKB/Swiss-Prot

Fungi
- UniProtKB/Swiss-Prot
- Norine toxic and non-toxic nonribosomal peptides
  (https://bioinfo.cristal.univ-lille.fr/norine/)

Bacteria
- UniProtKB/Swiss-Prot
- Norine

Viruses
- UniProtKB/Swiss-Prot
Protein toxins in UniProtKB/Swiss-Prot

- All: ~7000
- Animals: ~6300 (Tox-Prot)
- Plants: ~130
- Fungi: ~100
- Bacteria: ~440
- Viruses: ~60

www.uniprot.org

release 2020_05
Protein toxins in UniProtKB/Swiss-Prot

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
<th>Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>~7000</td>
<td>UniProtKB.keyword:&quot;Toxin [KW-0800]&quot; AND reviewed:yes</td>
</tr>
<tr>
<td>Animals</td>
<td>~6300</td>
<td>keyword:&quot;Toxin [KW-0800]&quot; taxonomy:&quot;Metazoa [33208]&quot;</td>
</tr>
<tr>
<td>Plants</td>
<td>~130</td>
<td>keyword:&quot;Toxin [KW-0800]&quot; taxonomy:&quot;Viridiplantae [33090]&quot;</td>
</tr>
<tr>
<td>Fungi</td>
<td>~100</td>
<td>keyword:&quot;Toxin [KW-0800]&quot; taxonomy:&quot;Fungi [4751]&quot;</td>
</tr>
<tr>
<td>Bacteria</td>
<td>~440</td>
<td>keyword:&quot;Toxin [KW-0800]&quot; taxonomy:&quot;Bacteria [2]&quot;</td>
</tr>
<tr>
<td>Viruses</td>
<td>~60</td>
<td>keyword:&quot;Toxin [KW-0800]&quot; taxonomy:&quot;Viruses [10239]&quot;</td>
</tr>
</tbody>
</table>
UniProt
UniProt Consortium

European Bioinformatics Institute
European Molecular Biology Laboratory

EMBL-EBI

SIB Swiss Institute of Bioinformatics

PIR Protein Information Resource

UniProt
UniProt – the Universal Protein resource

WELCOME

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

www.uniprot.org
UniProt – the Universal Protein resource
UniProt – the Universal Protein resource
Protein sequence
- The quality of the protein sequences is dependent on the information provided by the submitter of the original nucleotide entry (CDS) or of the gene prediction pipeline.

Biological information
Sources of annotation
- Provided by the submitter
- Automatic annotation from rules
UniProtKB/Swiss-Prot – Expert biocuration

Reviewed section

**Protein sequence** (combine together available coding sequences, annotate sequence discrepancies, report sequencing mistakes…)

**Biological information** (extract literature information, sequence analysis, ortholog data propagation, …)
What is inside a UniProtKB/Swiss-Prot entry?
Sequence features

Nomenclature and Taxonomy

Annotations

UniProtKB AC: P05484

Sequence

MKLTCVVIVA  VLLLTACQLI  TADDSRGTRQK  HRALRSTTKL  STSTFCKGKG

60  70

AKCSRLMYDC  CTGSCRSGKC  G
1 UniProtKB/Swiss-Prot entry

all protein products encoded by **ONE GENE** in a given species
Expert curation of sequences

An example from a King baboon spider toxin:

- one gene or two genes?
- sequencing error?
- variation on the same gene?
- taxonomical misattribution?

UniProtKB AC: D5J6X1
Expert curation of sequences

An example from a King baboon spider toxin:

- one gene or two genes?
- sequencing error?
- variation on the same gene?
- taxonomical misattribution?

Table 1 List of precursor sequences deduced from ESTs obtained from the cDNA library of C. cRAWs

<table>
<thead>
<tr>
<th>Feature key</th>
<th>Position(s)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence conflict</td>
<td>36</td>
<td>V → A in ADF28495 (PubMed:20372963)</td>
</tr>
</tbody>
</table>
Sequence features

Nomenclature and Taxonomy

Sequence

Annotations

---

**Molecule processing**

<table>
<thead>
<tr>
<th>Feature key</th>
<th>Position(s)</th>
<th>Description</th>
<th>Actions</th>
<th>Graphical view</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal peptide</td>
<td>1 – 22</td>
<td>Sequence analysis</td>
<td>Add BLAST</td>
<td></td>
</tr>
<tr>
<td>Propeptide</td>
<td>23 – 45</td>
<td>2 Publications</td>
<td>Add BLAST</td>
<td></td>
</tr>
<tr>
<td>Peptide</td>
<td>46 – 70</td>
<td>Omega-conotoxin MVIIA</td>
<td>Add BLAST</td>
<td></td>
</tr>
</tbody>
</table>

**Amino acid modifications**

<table>
<thead>
<tr>
<th>Feature key</th>
<th>Position(s)</th>
<th>Description</th>
<th>Actions</th>
<th>Graphical view</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disulfide bond</td>
<td>46 ↔ 61</td>
<td>7 Publications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disulfide bond</td>
<td>53 ↔ 65</td>
<td>7 Publications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disulfide bond</td>
<td>60 ↔ 70</td>
<td>7 Publications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified residue</td>
<td>70</td>
<td>Cysteine amide</td>
<td>1 Publication</td>
<td></td>
</tr>
</tbody>
</table>

UniProtKB AC: P05484
Nomenclature and Taxonomy

Sequence features

Sequence

Annotations

References

---

**Nomenclature and Taxonomy**

<table>
<thead>
<tr>
<th>Protein names</th>
<th>Recommended name: Omega-conotoxin MVIIA</th>
<th>1 Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Alternative name(s):</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• SNX-111</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• INN: Ziconotide</td>
<td>1 Publication</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organism</th>
<th>Conus magus (Magus cone) (Magician’s cone snail)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxonomic</td>
<td>6492 [NCBI]</td>
</tr>
<tr>
<td>identifier</td>
<td></td>
</tr>
<tr>
<td>Taxonomic</td>
<td>Eukaryota &gt; Metazoa &gt; Spiralia &gt; Lophotrochozoa &gt; Mollusca</td>
</tr>
<tr>
<td>lineage</td>
<td>Gastropoda &gt; Caenogastropoda &gt; Neogastropoda &gt; Conoidea &gt; Conidae &gt; Conus &gt; Pionoconus</td>
</tr>
</tbody>
</table>
Annotations

Sequence features

Nomenclature and Taxonomy

Sequence

Annotations

Function

Omega-conotoxins act at presynaptic membranes, they bind and block voltage-gated calcium channels. This toxin blocks Cav2.2/CACNA1B calcium channels (IC_{50}=180 nM). When injected in mammals, it induces adverse effects, such as tremor, diminution of spontaneous locomotor activity and bad coordinated locomotion (PubMed:26344359).

Pathology & Biotech

Toxic dose

LD_{50} is 0.10 mg/kg by intramuscular injection into goldfish (C.carassius). 1 Publication

Pharmaceutical use

Is available under the names Prialt by Neurex. It blocks acute pain in patients who no longer obtain relief from opiate drugs. It is 100 to 1000 times more potent than morphine. By blocking calcium channels it disable nerves that transmit pain signals. 1 Publication

Family & Domains

Domain

The presence of a 'disulfide through disulfide knot' structurally defines this protein as a knottin. 7 Publications

The cysteine framework is VI/VII (C-C-CC-C-C). Curated

Sequence similarities

Belongs to the conotoxin O1 superfamily. Curated
Annotations of enzymes

Sequence features

Nomenclature and Taxonomy

Sequence

SwissProt

Annotations

Catalytic activity

- an N-(acyl)-sphingosylphosphocholine = an N-(acyl)-sphingosyl-1,3-cyclic phosphate + choline

Source: Rhea. Show »

UniProtKB AC: P0CE80
Annotations of enzymes

Sequence features

Nomenclature and Taxonomy

Find Rhea reactions

Rhea is an expert-curated knowledgebase of chemical and transport reactions of biological interest - and the standard for enzyme and transporter annotation in UniProtKB. Rhea uses the chemical dictionary ChEBI (Chemical Entities of Biological Interest) to describe reaction participants. Learn more...

https://www.rhea-db.org/
Evidence tags

Function
Omega-conotoxins act at presynaptic membranes, they bind and block voltage-gated calcium channels. This toxin blocks Cav.2.2/CACNA1B calcium channels (IC_{50}=180 nM). When injected in mammals, it induces adverse effects, such as tremor, diminution of spontaneous locomotor activity and bad coordinated locomotion. [PubMed:26344359].

Pathology & Biotech

Toxic dose
LD_{50} is 0.10 mg/kg by intramuscular injection into rats.

Pharmaceutical use
Is available under the names Prialt by Neurex. It binds to calcium channels and disable nerves that transmit pain signals. It is 100 to 1000 times more potent than morphine. By blocking calcium channels via N-type calcium channel inhibitors, this toxin induces adverse effects, such as tremor, diminution of spontaneous locomotor activity and bad coordinated locomotion.

Expression

Tissue specificity
Expressed by the venom duct.

Family & Domains

Domain
The presence of a 'disulfide through disulfide knot' structurally defines this protein as a knottin.
The cysteine framework is VI/VII (C-C-C-C-C-C).

Sequence similarities
Belongs to the conotoxin O1 superfamily.
Ontologies

Sequence features

Nomenclature and Taxonomy

Sequence

Annotations

Ontologies

GO - Molecular function:
- ion channel inhibitor activity
- toxin activity

Complete GO annotation on QuickGO ...

GO - Biological process:
- pathogenesis

Complete GO annotation on QuickGO ...

Keywords:
- Calcium channel impairing toxin
- Ion channel impairing toxin
- Neurotoxin
- Presynaptic neurotoxin
- Toxin
- Voltage-gated calcium channel impairing toxin
### Links to other databases

#### Sequence databases
- **Select the link**
  - **EMBL**
  - **GenBank**
  - **DDBJ**
  - **PIR**

  **Destinations:**
  - **FJ959111** Genomic DNA Translation: **ADB93081.1**

  **3D structure databases**
  - **Select the link**
    - **PDB entry**
    - **Method:**
      - 1DW4: NMR
      - 1DW5: NMR
      - 1FEO: NMR
      - 1MVI: NMR
      - 1OMG: NMR
      - 1TT3: NMR
      - 1TTK: NMR

  **Organism-specific databases**
  - **ConoServer**
    - **1564**, MVIA precursor

#### Family and domain databases
- **InterPro**: View protein in InterPro
  - IPR004214, Conotoxin
  - IPR012321, Conotoxin_omega-typ_CS

- **Pfam**: View protein in Pfam
  - PF02950, Conotoxin, 1 hit

- **SITE**: View protein in PROSITE
  - PS60004, OMEGA_CONOTOXIN, 1 hit

---

[swissprot logo]
Helpdesk

help@uniprot.org

UniProtKB - P05484 (O17A_CONMA)

Protein: Omega-conotoxin MVIIA
Gene: N/A
Organism: Conus magus (Magus cone) (Magician's cone snail)
Function:
Omega-conotoxins act at presynaptic membranes, they bind and block voltage-gated calcium channels.
Summary

**UniProtKB/Swiss-Prot**

- is an expertly curated protein sequence knowledgebase
- contains sequences and functional information
- provides proteins from all kingdoms (animals, bacteria, viruses, plant, fungi)
- is highly cross-referenced to other databases
- is a freely available service for the scientific community
- public version is updated every two months
- Curation priorities to proteins studied at the functional level, or other information deemed important.
ConoServer
ConoServer is a database specializing in the sequence and structures of conopeptides, which are peptides expressed by carnivorous marine cone snails. A fascinating feature of these peptides is their high specificity and affinity towards human ion channels, receptors and transporters of the nervous system. This makes conopeptides an interesting resource for the physiological studies of neuroreceptors and promising drug leads. Conopeptides are further described here and a selection of recent reviews on the subject can be found here.

Classifications. Conopeptides are classified into disulfide rich (conotoxins) and several classes of disulfide poor peptides. The three classification schemes used in ConoServer, the gene superfamilies, the cysteine frameworks, and the pharmacological families are described and analyzed here.

Post-translational modifications. Conopeptides are heavily post-translationally modified and the list of modifications found naturally or introduced artificially is provided here.

Statistics. Statistics on the currently known conopeptides are provided here. The statistics include: the relationship between conopeptide classification schemes, the sequence consensus between signal peptides of each gene superfamily, the number of entries for each cone snail species, the characteristics of conopeptides for which a three-dimensional structure was determined, the number of patented sequences, and the journals the most cited in the ConoServer.

Provides information on cone snail toxins (conotoxins)

- Protein sequences
- Nucleotide sequences
- 3D structures
- Functions
- Tools (precursor, mass,...)
ConoServer Search

Protein Search

Permits to retrieve identical sequences (or part of them)
Not a prediction tool.
ConoServer is a database specializing in the sequences and structures of conopeptides, which are peptides expressed by carnivorous marine cone snails. A fascinating feature of these peptides is their high specificity and affinity towards human ion channels, receptors and transporters of the nervous system. This makes conopeptides an interesting resource for the physiological studies of neuroreceptors and promising drug leads. Conopeptides are further described here and a selection of recent reviews on the subject can be found here.

Classifications. Conopeptides are classified into disulfide rich (conotoxins) and several classes of disulfide poor peptides. The three classification schemes used in ConoServer, the gene superfamilies, the cysteine frameworks, and the pharmacological families are described and analyzed here.
ConoServer vs UniProtKB/Swiss-Prot

8069 entries
1 protein = several entries
1 entry/mature sequence
1 entry/precursor sequence
1 entry/synthetic constructs
All existing sequences known

1353 entries
keyword:"Toxin [KW-0800]"
taxonomy:"Conoidea [37797]"
1 entry = products of 1 gene
Only sequences with additional information (function, PTMs, 3D structure, …)
ArachnoServer
ArachnoServer

http://www.arachnoserver.org/mainMenu.html

Welcome to ArachnoServer

ArachnoServer is a manually curated database containing information on the sequence, three-dimensional structure, and biological activity of protein toxins derived from spider venom. Spiders are the largest group of venomous animals and they are predicted to contain by far the largest number of pharmacologically active peptide toxins (Escoubas et al., 2006). ArachnoServer has been custom-built so that a wide range of biological scientists, including neuroscientists, pharmacologists, and toxicologists, can readily access key data relevant to their discipline without being overwhelmed by extraneous information.

All spider toxin entries are sourced from UniProt/Swiss-Prot then manually curated by our expert team using available literature and patent information. Spider taxonomy is based on the latest version of the authoritative World Spider Catalog. A key feature of ArachnoServer is the use of a molecular target ontology based on the channel and receptor subtype definitions recommended by IUPHAR. Moreover, in addition to any legacy synonyms, all peptide toxins in the database have been assigned names according to the recently described rational nomenclature for spider toxins (King et al., 2005).

ArachnoServer allows advanced searches of toxin information, browsing, as well as similarity searches using BLAST. Each toxin record is displayed in a single page and, where available, a toxin's structure can be dynamically visualised.

You can get started by using the search box at the top of each page to search toxin names, synonyms, spider common names, and spider taxonomy. For advanced searches, click either the 'search' tab or the 'advanced' link below the search box. Help snippets are available at the top right corner of the 'search', 'browse' and 'blast' pages. You can download the user manual using the link at the top of each page.

Please find an example of curated toxin card here

Database Information

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toxins (total)</td>
<td>1576</td>
</tr>
<tr>
<td>Toxins (curated)</td>
<td>1458</td>
</tr>
<tr>
<td>Species</td>
<td>760</td>
</tr>
</tbody>
</table>

Please visit the download page to download data from ArachnoServer.

UPDATE: Please try out ToxNote, our new pipeline to significantly fast track the analysis of spider venom gland transcriptomes, by isolating all toxins, toxin-like sequences generated by large Next Generation (NG) sequencing projects.

Toxins deposited over time

Provide information on spider toxins:

- Protein sequences
- Nucleotide sequences
- 3D structures
- Functions
- Tools (precursor, toxin name generation, …)
ArachnoServer vs UniProtKB/Swiss-Prot

1458 entries

1 entry = 1 mature protein (shows precursors)
• protein centric

1431 entries
keyword: "Toxin [KW-0800]"
taxonomy: "Araneae [6893]"

1 entry = products of 1 gene
• gene centric
Kalium
<table>
<thead>
<tr>
<th>Organism</th>
<th>Name</th>
<th>Synor</th>
<th>UniProt ID</th>
<th>Sequin</th>
<th>PDB</th>
<th>Mass</th>
<th>Pub. Date</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bungarus fasciatus</td>
<td>Kunitz-type serine</td>
<td>Ku...</td>
<td>P25660</td>
<td>KNRPT1</td>
<td>1JC6</td>
<td>7288.15</td>
<td>1983</td>
<td>Kv1.3</td>
</tr>
<tr>
<td>Bungarus multicinctus</td>
<td>Kunitz-type serine</td>
<td>Ku...</td>
<td>Q1RPT0</td>
<td>RKRHFI</td>
<td></td>
<td>7122.03</td>
<td>2006</td>
<td></td>
</tr>
<tr>
<td>Bungarus multicinctus</td>
<td>Kunitz-type serine</td>
<td>Ku...</td>
<td>P00989</td>
<td>RKRHFI</td>
<td>1BUN</td>
<td>7191.14</td>
<td>1982</td>
<td>Kv1</td>
</tr>
<tr>
<td>Bungarus multicinctus</td>
<td>Kunitz-type serine</td>
<td>Ku...</td>
<td>P00987</td>
<td>RQRHRI</td>
<td></td>
<td>7182.09</td>
<td>1978</td>
<td>Kv1</td>
</tr>
<tr>
<td>Crotalus adamanteus</td>
<td>Myotoxin</td>
<td>My...</td>
<td>P24330</td>
<td>YKRCCH</td>
<td></td>
<td>5509.48</td>
<td>1991</td>
<td></td>
</tr>
<tr>
<td>Crotalus durissus cuman...</td>
<td>Crotamine-IV-3</td>
<td>Cro...</td>
<td>P86194</td>
<td>YKRCCH</td>
<td></td>
<td>4951.88</td>
<td>2007</td>
<td></td>
</tr>
<tr>
<td>Crotalus durissus cuman...</td>
<td>Crotamine-IV-2</td>
<td>Cro...</td>
<td>P86193</td>
<td>YKRCCH</td>
<td></td>
<td>4900.87</td>
<td>2007</td>
<td></td>
</tr>
<tr>
<td>Crotalus durissus ruruima</td>
<td>Crotamine Ile-19</td>
<td>Cro...</td>
<td>P63327</td>
<td>YKRCCH</td>
<td>4GV5</td>
<td>4883.71</td>
<td>1993</td>
<td></td>
</tr>
<tr>
<td>Crotalus durissus terrificus</td>
<td>Crotamine</td>
<td>Cro...</td>
<td>Q9PW6F5</td>
<td>YKRCCH</td>
<td>1H5O,1</td>
<td>4883.71</td>
<td>1975</td>
<td>Kv1.1,Kv1.2,Kv1</td>
</tr>
<tr>
<td>Crotalus oreganus conc...</td>
<td>Myotoxin-2</td>
<td>My...</td>
<td>P12029</td>
<td>YKRCCH</td>
<td></td>
<td>5027.89</td>
<td>1987</td>
<td></td>
</tr>
<tr>
<td>Pub. Date</td>
<td>Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1983</td>
<td>Kv1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1982</td>
<td>Kv1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1978</td>
<td>Kv1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975</td>
<td>Kv1.1,Kv1.2,Kv1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Crotamine

<table>
<thead>
<tr>
<th>Permanent Link</th>
<th><a href="https://kaliumdb.org/toxins/338">https://kaliumdb.org/toxins/338</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism</td>
<td>Crotalus durissus terrificus</td>
</tr>
<tr>
<td>Name</td>
<td>Crotamine</td>
</tr>
<tr>
<td>Synonyms</td>
<td>Crotamine; Crt; Myotoxin</td>
</tr>
<tr>
<td>UniProt ID</td>
<td>Q9PWF3</td>
</tr>
<tr>
<td>Sequence</td>
<td>YKQCHHKGGHCPFPEKIKCLPPSSDFGKMDGRWRWKCCKGSG</td>
</tr>
<tr>
<td>Raw Sequence</td>
<td>MKILYLLFAFLFLAFALSEPGNAYYKQCHHKGGHCPFPEKIKCLPPSSDFGKMDGRWRWKCCKGSGK</td>
</tr>
<tr>
<td>Mode</td>
<td>blocker</td>
</tr>
<tr>
<td>PDB</td>
<td>1H5O, 1Z99</td>
</tr>
<tr>
<td>Mass</td>
<td>4883.71</td>
</tr>
<tr>
<td>Publication Date</td>
<td>1975</td>
</tr>
<tr>
<td>Last Modified</td>
<td>2017-11-22</td>
</tr>
</tbody>
</table>

#### Activity measures

<table>
<thead>
<tr>
<th>Target Channel</th>
<th>Data Type</th>
<th>Method Type</th>
<th>Test System</th>
<th>Species</th>
<th>Value, nM</th>
<th>Ref. (PubMed/DOI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kv1.1</td>
<td>IC50</td>
<td>volt</td>
<td>oocyte</td>
<td>rat</td>
<td>369</td>
<td>PM: 22498659</td>
</tr>
<tr>
<td>Kv1.2</td>
<td>IC50</td>
<td>volt</td>
<td>oocyte</td>
<td>rat</td>
<td>386</td>
<td>PM: 22498659</td>
</tr>
</tbody>
</table>
Protein toxin databases
  Introduction
  UniProtKB/Swiss-Prot
  ConoServer
  ArachnoServer
  Kalium

Other resources on protein toxins
  Toxin prediction tools
  VenomZone
  Knottin
Toxin prediction tools
# Toxin prediction tools

<table>
<thead>
<tr>
<th>Predictor name</th>
<th>Predicts</th>
<th>Organisms</th>
<th>Last update</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTX pred</td>
<td>Neurotoxin</td>
<td>Animals and Bacteria</td>
<td>2007 (creation)</td>
</tr>
<tr>
<td><a href="http://crdd.osdd.net/raghava/ntxpred/">http://crdd.osdd.net/raghava/ntxpred/</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTXpred</td>
<td>Toxin</td>
<td>Bacteria</td>
<td>2007 (creation)</td>
</tr>
<tr>
<td><a href="http://crdd.osdd.net/raghava/btxpred/">http://crdd.osdd.net/raghava/btxpred/</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ClanTox</td>
<td>Toxin</td>
<td>Animals</td>
<td>2008</td>
</tr>
<tr>
<td><a href="http://www.clantox.cs.huji.ac.il/about.php">http://www.clantox.cs.huji.ac.il/about.php</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ToxinPred</td>
<td>Short toxin (≤35 aa)</td>
<td>Animals and Bacteria</td>
<td>2013 (creation)</td>
</tr>
<tr>
<td><a href="https://webs.iiitd.edu.in/raghava/toxinpred/index.html">https://webs.iiitd.edu.in/raghava/toxinpred/index.html</a></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**UniProtKB/Swiss-Prot**
- Release 54.0 of 24 July 2007
  - 1,899 animal toxins
- Release 2020_05 of Oct-07, 2020
  - 6,322 animal toxins
  - x 3.3
VenomZone
VenomZone is a free web resource that provides information on venoms from six taxa (snakes, scorpions, spiders, cone snails, sea anemones and insects), as well as on their targets. Information can be browsed through pages on taxonomy, activity and venom protein families and all these pages link to related venom protein information from the manually curated UniProtKB/Swiss-Prot database.
VenomZone is a free web resource that provides information on venoms from six taxa (snakes, scorpions, spiders, cone snails, sea anemones and insects), as well as on their targets. Information can be browsed through pages on taxonomy, activity and venom protein families and all these pages link to related venom protein information from the manually curated UniProtKB/Swiss-Prot database.
### Pharmacological target of venom proteins

This page gives the list of channel/receptor families that are targeted by venom proteins and an access to these venom proteins. A detailed information on these families is available by clicking on the “channel/receptor family” blue hyperlink, while clicking on the logo permits to retrieve venom protein(s) of interest in UniProtKB/Swiss-Prot. Dashes indicate that no venom protein has been described yet to target the channel/receptor family.

The column “Channel/Receptor Human” of the table permits an access to human channel/receptor entries of UniProtKB/Swiss-Prot.

Another view of this table is also available in Animal toxin annotation program, Statistics.

See also Venom protein activity
See also Systems targeted by venom proteins
See also Activity Greek symbol descriptors

<table>
<thead>
<tr>
<th>Channel/receptor families</th>
<th>Voltage-gated sodium (NaV) channel</th>
<th>Voltage-gated potassium (Kv) channel</th>
<th>Calcium-activated potassium (KCa) channel</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-protein coupled receptor (GPCR)</td>
<td>Human</td>
<td>Snake</td>
<td>Scorpion</td>
</tr>
<tr>
<td>Adrenergic receptor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bradykinin receptor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cholecystokinin receptor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receptor TM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epidermal growth factor receptor (EGFR)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insulin receptor (INSR)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vascular endothelial growth factor receptor (VEGFR)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transporter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glutamate transporter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurotransmitter transporter</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Voltage-gated potassium channel impairing toxin

### Channel
Voltage-gated potassium channels (Kv) are composed of four transmembrane subunits; each is analogous to a single domain of the principal subunits of sodium or calcium channels. Kv are remarkable for their diversity. They include 40 different channels that are classified into 12 distinct subfamilies based on their amino acid sequence homology (Kv1 to Kv12). These subunits can assemble into homo- and hetero-tetrmers, leading to a wide diversity of different channel complexes. The diversity of potassium channels allows neurons and other excitable cells to precisely tune their electrical signaling properties by expression of different combinations of potassium channel subunits.

Kv are activated by depolarization, and the outward movement of potassium ions through them repolarizes the membrane potential to end action potentials, hyperpolarizes the membrane potential immediately following action potentials, and plays a key role in setting the resting membrane potential. In this way, potassium channels control electrical signaling in excitable cells and regulate ion flux and calcium transients in nonexcitable cells.

### Toxin

<table>
<thead>
<tr>
<th>Kv</th>
<th>Toxins from different taxa that target Kv channels.</th>
<th>Click on to retrieve toxins of interest.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Link**
- Back to Pharmacological target
- Back to Ion channel impairing toxin

**Principal activity**
Neurotoxin

**Nomenclature**
Toxins that inhibit or block Kv channels are named k toxins (see King et al., 2008).

In scorpion, potassium channel toxins (KTx) have been classified and the different sub-families are described and listed here.

**Reference**
- Review: Catterall et al., 2007
### UniProtKB 2020_05 results

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

### Popular organisms

<table>
<thead>
<tr>
<th>Organism</th>
<th>Entry name</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mesobuthus martensli (Manchurian scorpion)</td>
<td>KAX36_MESMA</td>
<td>60</td>
</tr>
<tr>
<td>(Buthus martensi)</td>
<td>(Manchurian scorpion)</td>
<td></td>
</tr>
<tr>
<td>(Buthus martensi)</td>
<td>(Buthus martensi)</td>
<td></td>
</tr>
<tr>
<td>Scorpion palmaus (Israeli golden scorpion)</td>
<td>KAX62_SCOPA</td>
<td>34</td>
</tr>
<tr>
<td>(Scorpio maurus palmaus)</td>
<td>(Scorpio maurus palmaus)</td>
<td></td>
</tr>
<tr>
<td>Hoffmanihadrurus gertschi (Scorpion)</td>
<td>KBX3_HOFGE</td>
<td>95</td>
</tr>
<tr>
<td>(Hoffmannihadrurus gertschi)</td>
<td>(Hoffmannihadrurus gertschi)</td>
<td></td>
</tr>
</tbody>
</table>
VenomZone is a free web resource that provides information on venoms from six taxa (snakes, scorpions, spiders, cone snails, sea anemones and insects), as well as on their targets. Information can be browsed through pages on taxonomy, activity and venom protein families and all these pages link to related venom protein information from the manually curated UniProtKB/Swiss-Prot database.
# Venom protein family

This page gives a list of venom protein families that are well described in the literature or found in different taxonomic groups. Those that are specific to a venomous taxa are marked by an asterisk. A detailed information of these families is available by clicking on the "protein family" blue hyperlink.

To have an idea of which taxon is known to belong to a protein family, logos or dashes are indicated in the table. Logos show that at least one venom protein belongs to the family, while dashes indicate that no venom protein has been described yet to belong to the family. Logos are clickable and provide a direct access to related UniProtKB/Swiss-Prot entries.

<table>
<thead>
<tr>
<th>Protein families</th>
<th>Snake</th>
<th>Scorpion</th>
<th>Cone snail</th>
<th>Sp</th>
</tr>
</thead>
<tbody>
<tr>
<td>(lysosomal acid and phospholipase A1)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Arthropod Ce/INH/IG/IV/IR hormone</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Arthropod dermonecrotic toxin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AVIT (prolincin)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bradykinin-related peptide</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bradykinin-potentiating peptide</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Catterall peptidase</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cathepsin E protein (CERP)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Complement C5b-9 complex</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CRISP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Crotamine-myrteoxin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cystatin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Destrin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>EGF domain peptide</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Endothelin/endothelin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Fvain meningococcus oxidase (L-amino acid oxidase)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Glyceraldehyde-3-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Glyceraldehyde-2-2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Insulin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Metalloprotease</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Neutral protease</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NOD-12</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nucleotide pyrophosphatase/phosphodiesterase</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Oligopeptide</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PEGF/VEGF growth factor</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Peptidase B1 (serine protease)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Phospholipase A2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Phospholipase B-like</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Knottin
The Knottin database and tool

https://www.dsimb.inserm.fr/KNOTTIN/

Welcome to the KNOTTIN database

What are knottins?

- Knottins are small disulfide-rich proteins characterized by a very special “disulfide through disulfide knot”
- This knot is achieved when one disulfide bridge crosses the macrocycle formed by the two other disulfides and the interconnecting backbone.
- The knot implies that knottins contain at least 3 disulfide bridges.
- The structural family of knottins have the disulfide between cysteines III and VI (blue) going through disulfides I-V and II-V (green).
- The growth factor cystine knots also contain a knot but the connectivity is different and they cannot be superimposed onto knottins. These proteins belong to a distinct structural family not described in this site.
- Knottins are sometime refered to as "Inhibitor Cystine Knots".

Disulfide connectivity

Knottins  Growth factor cystine knots

Updates

- June 2017
  Major update of the KNOTTIN database
  + Addition of theoretical 3D models
  + Visualization with JSmol
  + New design of the web interface

- February 2014
  The KNOTTIN database is available
  We had to stop the server for technical reasons. It is now available again.
  Please let us know about any trouble you could encounter with this server.
  Also note that we are currently working on the next update of the database.
  Thank you for using KNOTTIN!
In UniProtKB/Swiss-Prot (release 2020_05)

~80% of knottins are toxins
~20% are not toxins
Summary

- Generalist database UniProtKB/Swiss-Prot
- Organism- and target-specific db: ConoServer, ArachnoServer, Kalium
- All data can be used for machine learning (prediction tools)
- VenomZone portal (animal toxin data access through multiple points of view)
Pls: Alex Bateman, Alan Bridge, Cathy Wu

Key staff: Cecilia Arighi (Curation), Lionel Breuza (Curation), Elisabeth Coudert (Curation), Hongzhan Huang (Development), Damien Lieberherr (Curation), Michele Magrane (Curation), Maria Martin (Development), Peter McGarvey (Content), Darren Natale (Content), Sandra Orchard (Content), Ivo Peduzzi (Curation), Sylvain Poux (Curation), Manuela Pruess (Coordination), Shriya Raj (Coordination), Nicole Redaschi (Development)


European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, UK
Protein Information Resource (PIR), Washington DC and Delaware, USA
SIB Swiss Institute of Bioinformatics (SIB), Geneva, Switzerland
A particular thanks to

My colleagues
Andrea Auchincloss (expert in bacteria)
Anne Estreicher, Marie-Claude Blatter (expert in mammals)
Marc Feuermann (expert in fungi)
Damien Lieberherr (expert in plants)

My hierarchy
Elisabeth Coudert
Alan Bridge (Swiss-Prot group director)
UniProt funders

Swiss Federal Government through the State Secretariat for Education, Research and Innovation (SERI);

National Cancer Institute; National Eye Institute; National Human Genome Research Institute; National Heart, Lung, and Blood Institute; National Institute on Aging [U24HG007822]; National Institute of Allergy and Infectious Diseases; National Institute of Diabetes and Digestive and Kidney Diseases; National Institute of General Medical Sciences; National Institute of Mental Health of the National Institutes of Health [U24HG007822]; National Human Genome Research Institute [U41HG007822, U41HG002273]; National Institute of General Medical Sciences [R01GM080646, P20GM103446, U01GM120953];

European Molecular Biology Laboratory core funds; Biotechnology and Biological Sciences Research Council [BB/M011674/1]; British Heart Foundation [RG/13/5/30112]; Open Targets.
Thank you for your attention