ArachnoServer
SPIDER TOXIN DATABASE

ArachnoServer Manual

The University Of Queensland AUSTRALIA | IMB Institute for Molecular Bioscience

In collaboration with

QFAB DRIVING YOUR RESEARCH FURTHER

List of Figures

List of Tables

Introduction

ArachnoServer provides a single source of high-quality manually-curated information about proteinaceous spider toxins. The database has been designed specifically for the research needs of a wide rage of biological scientists, including pharmacologists, neuroscientists, toxinologists, medicinal chemists, ion channel scientists, clinicians, and structural biologists.

At the time of writing, ArachnoServer makes available 748 manually curated spider toxin records. As more spider toxins are discovered or known toxins further characterised, the curation team carefully collates this additional information into the database. As can be seen in Figure 1, the rate of discovery of spider toxins is increasing exponentially.

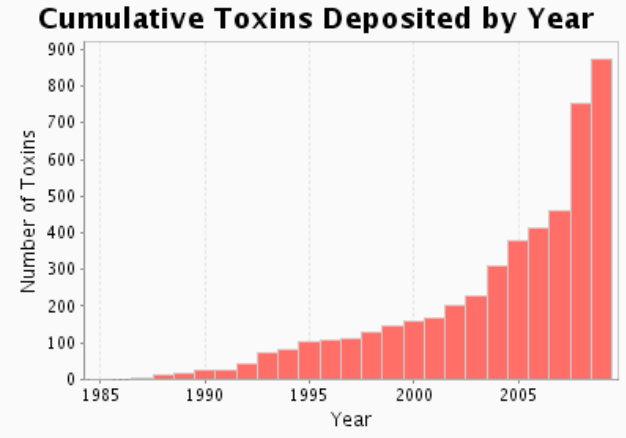


Figure 1. Cumulative number of spider toxins deposited into public databases since 1984.

We are committed to maintaining the most up-to-date and accurate repository of spider toxin information for the scientific community. We use ArachnoServer in our daily research, and we hope you find it useful too. Should you have any questions, comments, or feedback, please contact us at support@qfab.org.

What data is stored in ArachnoServer and how is it curated?

The starting point for data curation is the automated collection of all publically available sequence and annotation information for spider toxins from UniprotKB/SwissProt, INSDC (The International Nucleotide Sequence Database Collection; DDBJ, EMBL and GenBank) and the Protein Data Bank (PDB). We join these data sets with the assistance of Sequence Retrieval System (SRS) into a single non-redundant set, typically containing peptide sequences, nucleotide (mRNA) sequences, and protein structures (when available) for each individual toxin. During this process, we record all other database identifiers (for example NCBI taxonomy codes, Gene Ontology classifications, PROSITE and Pfam accessions and others; see the Database accessions section for more information) in the ArachnoServer database. We also keep records of all literature references, sequence and structure feature annotations (for example known locations of disulfide bonds) and descriptions of toxins.

ArachnoServer stores additional data sets (summarized in Figure 2). These include:

- The World Spider Catalog.
- NCBI Taxonomy (for association of a toxin's phyletic specificity to a consistent taxonomic ontology).
- A molecular target ontology we specifically created based on the channel and receptor subtype definitions and nomenclature recommended by the International Union of Basic & Clinical Pharmacology (IUPHAR).
- Lists of post-translational modifications (sourced from ConoServer with some additions), biological activities, sequence features, spider common names and even units used to describe LD_{50}, PD_{50}, and IC_{50} values.
- A repository of high-resolution photos for spiders with toxins in the database.

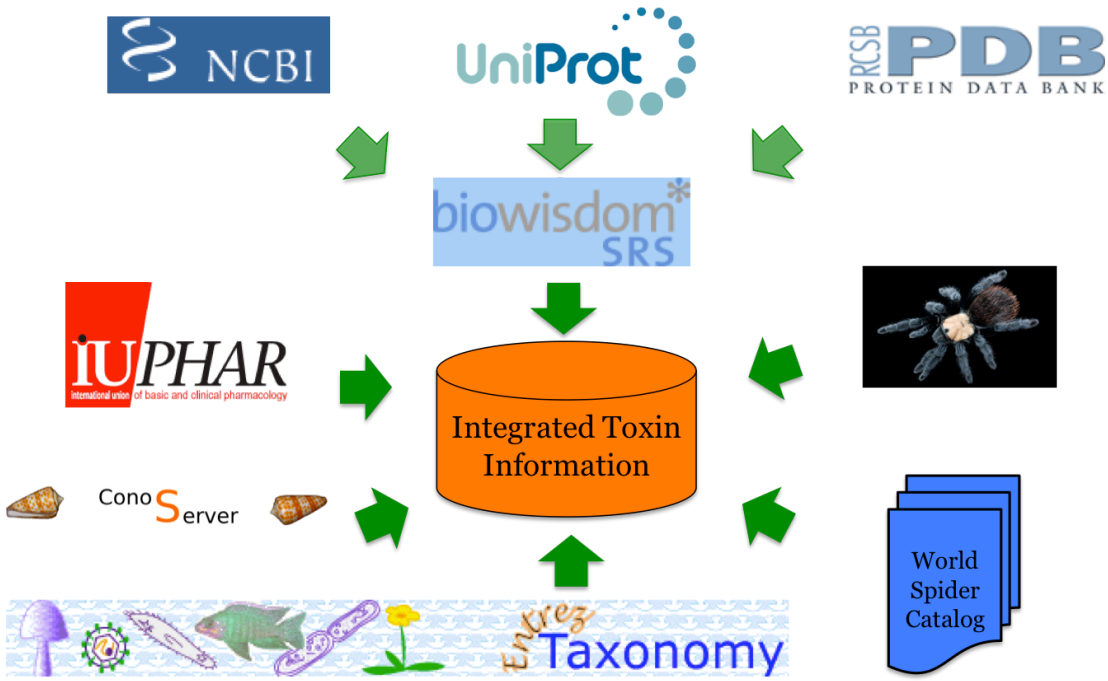


Figure 2. Data sets integrated into ArachnoServer.

With the publicly available toxin information, and the curated data sets described above, the curation team then work through each toxin from within the curation interface, ascribing a toxin name, and exhaustively collecting and summarising information from literature searches and patent information with data from public databases. Taxonomic source species for toxins in the public data set are matched, or corrected, with the current species from the World Spider Catalog, and annotated with a unique Life Science Identifier (LSID). High-resolution photos are added to spider species if available. Names for peptide toxins (defined as those smaller than 10 kDa) are ascribed using the rational nomenclature described by King *et al.* (2008), which is summarized at http://www.venomics.org/nomenclature, while sphingomyelinase nomenclature is derived from Binford *et al.* (2009). All known synonyms are included as well.

Once a toxin is curated, it is published to the public interface and then becomes searchable. Toxin's can, at any stage, be added or further curated by the team as additional information becomes available.

Searching ArachnoServer 

ArachnoServer supports both basic and advanced searches. In the banner at the top of every page is the basic search field (Figure 3). Search terms entered here query toxin names (including synonyms), taxonomic information (including historic taxonomy), and common names of spiders. If no search term is entered, and the search button is clicked, the search will return all toxins in the database.

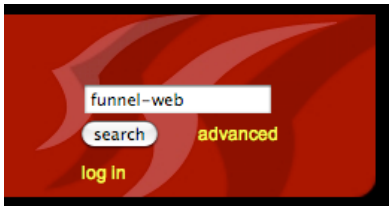


Figure 3. The basic search field, which is available on the banner of every page.

Advanced searches

For more specific queries, ArachnoServer provides an advanced search interface. Context-specific fields are dynamically arranged on the page, and search keys populated in drop-down lists of relevant data from the database. Search clauses can be grouped, then joined using boolean operators (such as 'or' and 'and'). A large selection of database fields can be searched, including all basic search fields, as well as additional features such as biological activity, posttranslational modifications, literature references, and counts of various data fields (e.g., the number of disulfide bonds in a toxin). The figures below show an advanced search being constructed. In this search we are particularly interested in finding toxins that have antimicrobial activity.

Select 'Biological Activities' ➜ 'include' ➜ 'Antimicrobial', then click the 'Search' button. This returns 40 toxins from the database, although only the first four toxins are displayed for brevity in Figure 4.

Search Toxins help

Biological Activities | include | Antimicrobial | grouping | +

Search | Reset

40 toxins found, displaying 1 to 15.

[First/Prev] **1**, 2, 3 [Next/Last]

Name	Synonym	Genus	Species
M-ctenitoxin-Cs1a	Cupiennin 1a	Cupiennius	salei
M-ctenitoxin-Cs1b	Cupiennin 1b	Cupiennius	salei
M-ctenitoxin-Cs1c	Cupiennin 1c	Cupiennius	salei
M-ctenitoxin-Cs1d	Cupiennin 1d	Cupiennius	salei

Figure 4. Using the advanced search interface to search for toxins with antimicrobial activity.

We can further limit the list of toxins by refining the search to include only those with known activity on *Escherichia coli*. From the search interface click the (+) (plus) button to create a new search clause, and then choose 'And' ➜ 'Phyletic Specificities' ➜ 'include' ➜ 'Escherichia coli'. Then click the 'Search' button. This time the search returns only 11 toxins, the first three of which are shown in Figure 5..

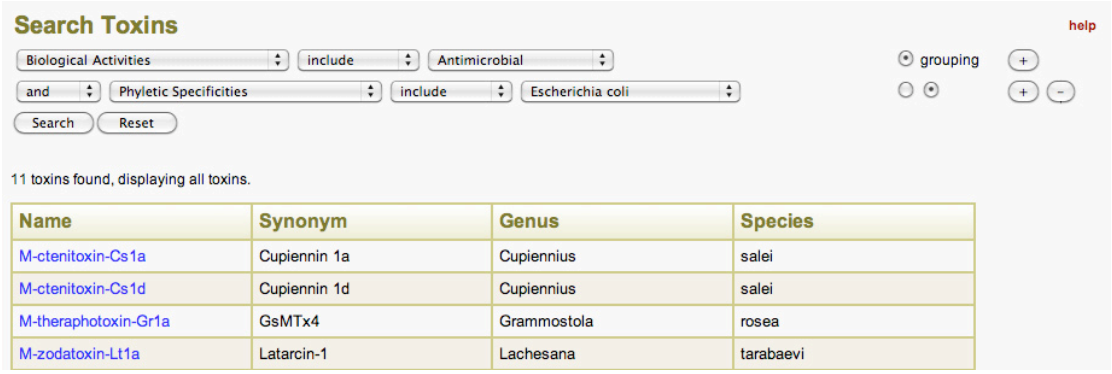


Figure 5. The results of adding an extra search term to the advanced search query to limit the results to only those toxins with antimicrobial biological activity against *Escherichia coli*.

We can further limit the number of toxins by any number of criteria. For example a medicinal chemist may be interested in only those toxins with a solved structure. To find this subset of toxins, add an extra search clause, choose 'And' ➜ 'Number of solved PDB structures' ➜ 'greater than or equal to (>=) -> '1', then click the 'Search' button. This limits the results to only three toxins (Figure 6).

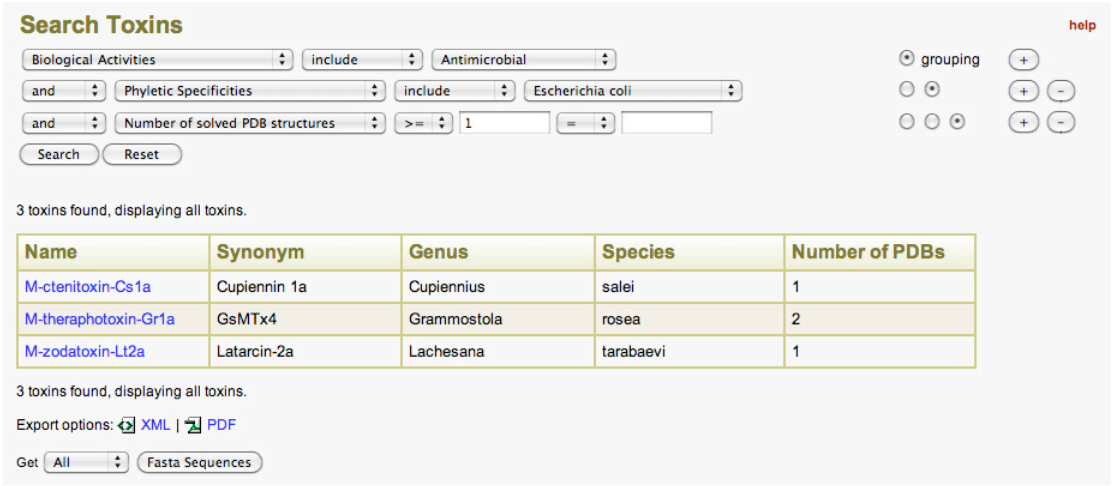


Figure 6. The results of adding an extra search term to the advanced search query that further limits the results to show only toxins with a solved PDB structure.

The advanced search interface is extremely flexible. For example, with the above query, we could change the species from *Escherichia coli* to *Arthrobacter globiformis*, click search, and find that a single toxin has antimicrobial activity on *Arthrobacter globiformis* as well as a solved PDB structure. Alternatively we might be interested not in antimicrobial toxins with solved structures, but those discovered at a particular time, for example the year 2006. We can do this by changing the 'number of solved PDB

structures' to 'Toxin Discovery Year' then choosing 'equals (=)' ➜ 2006 ➜ 'Search'. In this case ArachnoServer returns 8 toxins.

Advanced search fields

Table 1 lists searchable fields and provides a brief description of each.

Table 1. Database fields searchable using ArachnoServer's advanced search interface.

Field Category	Field	Data Type	Description
Toxin Metadata	Toxin Name	Text	Free text to search all the recommended toxin names in the database. Toxin names are ascribed using the rational nomenclature described in King et al, 2008.
	Toxin Description	Text	Free text describing the toxin.
	Generic Toxin Group	List	Toxins are grouped by taxonomic family. The toxin group is used when naming the toxin (see King et al, 2008 and http://www.venomics.org/nomenclature/).
	Toxin Discovery Year	Number	The date the toxin was discovered (note this has been carefully curated, and may differ from the date the toxin was originally deposited in the public database).
	Toxin Synonym	Text	ArachnoServer keeps a record of all names used to describe toxins (including abbreviations), and these can be searched using this field.
	Toxin Accession Number	Text	Use this field to search for a toxin by its accession number in one of the INSDC databases or other data repository such as Gene Ontology or InterPro
	Number of solved PDB structures	Number	Use this field to search for toxins with zero, one or any number of solved structures submitted to PDB.
	Date Toxin Added	Date	Date a toxin was submitted to UniProtKB/Swiss-Prot
Biological Activity	Biological Activities	List	Search for toxins with a specified biological activity.
	Phyletic Specificities	List	Search for toxins known to act on a particular species. The list of species is sourced from the NCBI Taxonomy.
	Phyla Effective Dosage (ED50), Lethal Dosage (LD50) and Paralytic Dosage (PD50)	Number, pmol/g	When known, search for toxins with specific ED50, LD50 and PD50 in consistent units of pmol/g. All literature units have been converted to pmol/g to allow for easy comparison of toxins.
	Number of Biological Activities	Number	Use this field to search for toxins with a specific number of known biological activities
	Number of Molecular Targets	Number	Search for toxins with a specific number of molecular targets.
Protein Information	Posttranslational modifications	List	Search for toxins with a specific type of posttranslational modification.
	Number of Posttranslational modifications	Number	Search for toxins with a specific number of posttranslational modifications.
	Number of Disulfide Bonds	Number	Search according to the number of disulfide bonds in a toxin (regardless of whether they are classified as experimentally determined, predicted or by homology to another toxin).
Source Species Taxonomy	Species	List	Search the toxins according to the species of spider from which they derive. Species names are sourced from the World Spider Catalog. Species with an asterisk (*) next to them also have downloadable photos available.
	Family	List	Search the database for toxins from spiders belonging to a specific taxonomic family.
	Common Name	List	Search the database according to common spider names.
	Historic Taxonomy	Text	Many spider species have been taxonomically reclassified in the past and this process is ongoing. ArachnoServer uses historic taxonomy data from the World Spider Catalog to make it possible to search for toxins using outdated taxonomy.

Literature Information	Journal Article Title	Text	Search for a toxin according to words in the title of citations in the data record for that toxin.
	Author List	Text	All authors of journal citations are stored in ArachnoServer. You can search for toxins published by a particular author using this field.
	Year	Number	Search for all toxins described in journal articles by the year of article publication, or range of years.
	Deposition Authors	Text	ArachnoServer also stores the names of those who originally deposited the toxin sequence in UniProtKB/Swiss-Prot, or the original INSDC database.
	Patent ID	Text	In cases where a toxin has been patented, and this information is publicly available, you can search using the patent ID and this field.
Mature Toxin Mass	Average Reduced mass, Average Oxidised mass, Monoisotopic Reduced mass and Monoisotopic Oxidised mass	Number	Search based on the mass of the mature toxin in daltons (Da). Most databases provide the mass of the reduced form of proteins. However, venom proteins are often rich in cysteines and these cysteines are invariably found in the oxidized form as disulfide bonds. Thus, for venom toxins, the oxidized mass is usually more relevant. ArachnoServer thus provides the mass of both the fully reduced and fully oxidized forms of the toxin. Moreover, ArachnoServer also provides the monoisotopic mass (which is used in mass spectrometry) as well as the more commonly used average mass. The advanced search enables searches using any of these mass categories.

Negating search terms, and specifying numeric ranges

We have tried to make the advanced search as intuitive as possible by providing an interface that closely resembles natural language, while being specific enough to drill down into the database quickly and accurately. For example, if text can be used to specify a search term, we provide 'like', 'not like', 'is' and 'is not', where terms specified in conjunction with 'like' and 'not like' are wrapped with wild card characters. Numeric data can be specified using a range with 'equals', 'not equals', 'greater than or equal to', and 'less than or equal to'. Note that only the first field is required for numeric data, but a full range can be specified using both available fields.

Joining search clauses using boolean operators, and grouping search clauses

The advanced search interface allows for up to 6 fields to be joined in a single search clause using boolean operators. To add a search field, press the  plus button on the far right of the screen. Search clauses are joined with one of three boolean terms: 'and', 'or' and 'but not'.

Search clauses can also be grouped together. Groupings can be specified using the radio buttons at the right hand side of the search clauses and look like this:

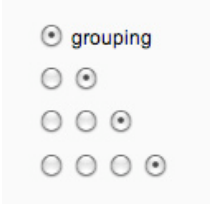


Figure 7. Radio buttons used to group search clauses.

Groups are defined vertically in the radio buttons. By default each search clause is placed into its own group (as in
Figure 7). In such a case, each clause will be joined with the clause immediately preceding it *and* immediately following it (if those clauses exist). If you would like to change this behaviour, you can choose to group the clauses together. Consider the example in Figure 8:

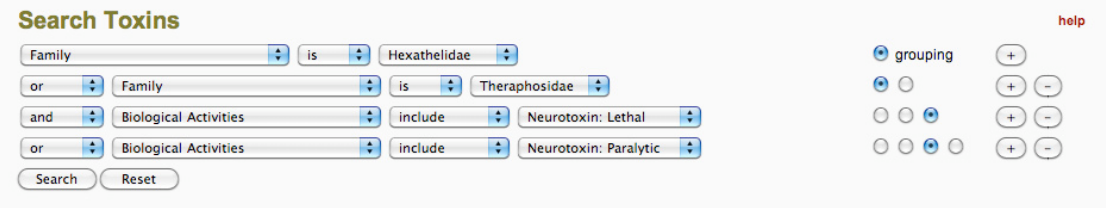


Figure 8. An example of joining search clauses together using the 'grouping' radio buttons.

This search is asking 'show me all toxins from the families Hexathelidae or Theraphosidae which display a biological activity of either *neurotoxin (lethal)* or *neurotoxin (paralytic).* The four search clauses are grouped together into two groups, the first group specifies the Araneae families, and the second the biological activities. It's easy to see that if we changed the second 'or' to 'and', then this search would return only those toxins in either family with both *neurotoxin (lethal)* and *neurotoxin (paralytic)* biological activities. If we changed the first 'or' to 'and', then this search would return zero results as expected (as species cannot belong to more than one taxonomic family). Many combinations of these search terms are possible.

Exporting search results, and toxin sequences in FASTA format

Search results are returned in a paginated table. Only general information is displayed for each toxin within the search results table to minimise clutter on the screen. At the bottom of each search result page are export options (**Error! Reference source not found.**).

Figure 9. Export options available for search results.

Data can be exported in both PDF and XML formats. Exporting data from any result page (for a multi-page set of search results) will export all data in the current search results. Additional information about each toxin is provided in the PDF export (description, synonyms, and taxonomic family). XML exported data is designed to be reformatted programmatically and may contain some HTML embedded codes (for example '<' and '>' characters are encoded as '>' and '<'), whereas PDF exported data is designed to

be readable in report format. FASTA sequences can be exported for all molecule types, or for specific types (peptide, DNA or mRNA).

Browsing ArachnoServer 

The *Browse* tab is always available at the top of the page. Toxins can be browsed by any of four categories: *Araneae taxonomy*, *Molecular Targets*, *Posttranslational Modifications* and *Phyletic Specificity*. Each category creates a different browsing tree on the right hand side of the screen for easy selection of toxins. Figure 10 is a screenshot of the tree that is visible when browsing ArachnoServer according to Araneae (spider) taxonomy. As with all browse trees, the taxonomy tree only contains species for which there are corresponding toxins in the database. As can be seen, selecting a family of spiders will display a table of all toxins characterised from a genus within that family (in this case the genus *Apomastus* within the family Cyrtaucheniidae). FASTA sequences can be exported from the *Browse* results, and clicking on the toxin name will open the Toxin Card for that particular toxin within a new browser tab (see the section below entitled Each toxin has its own 'Toxin Card' for more information).

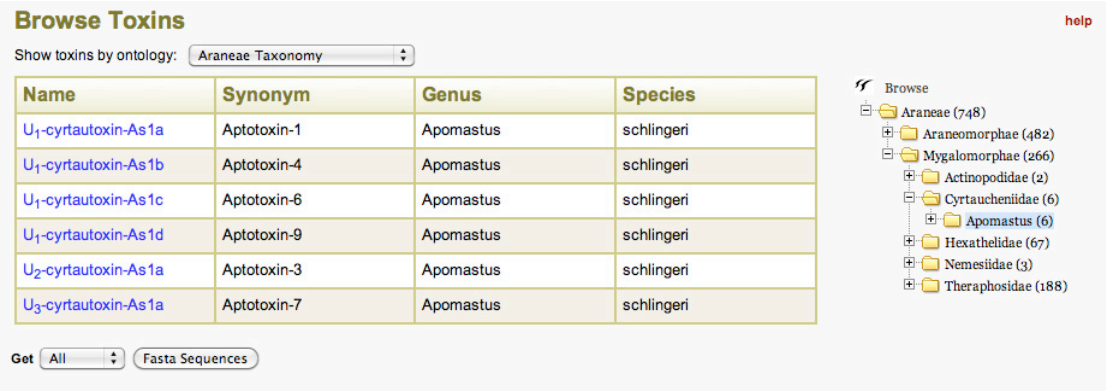


Figure 10. Browsing ArachnoServer by Araneae taxonomy.

Choosing a different ontology will reset the browse page and create a fresh browsing tree for the chosen ontology.

Browsing using the molecular target ontology is a convenient and powerful way to find all toxins that target, for example, a particular type or subtype of ion channel or receptor. Figure 11 shows the result of browsing using the molecular target ontology to search for toxins that specifically target invertebrate voltage-gated calcium channels. The molecular target ontology, which was developed specifically for ArachnoServer, is based on the channel and receptor subtype definitions and nomenclature recommended by the International Union of Basic & Clinical Pharmacology (IUPHAR).

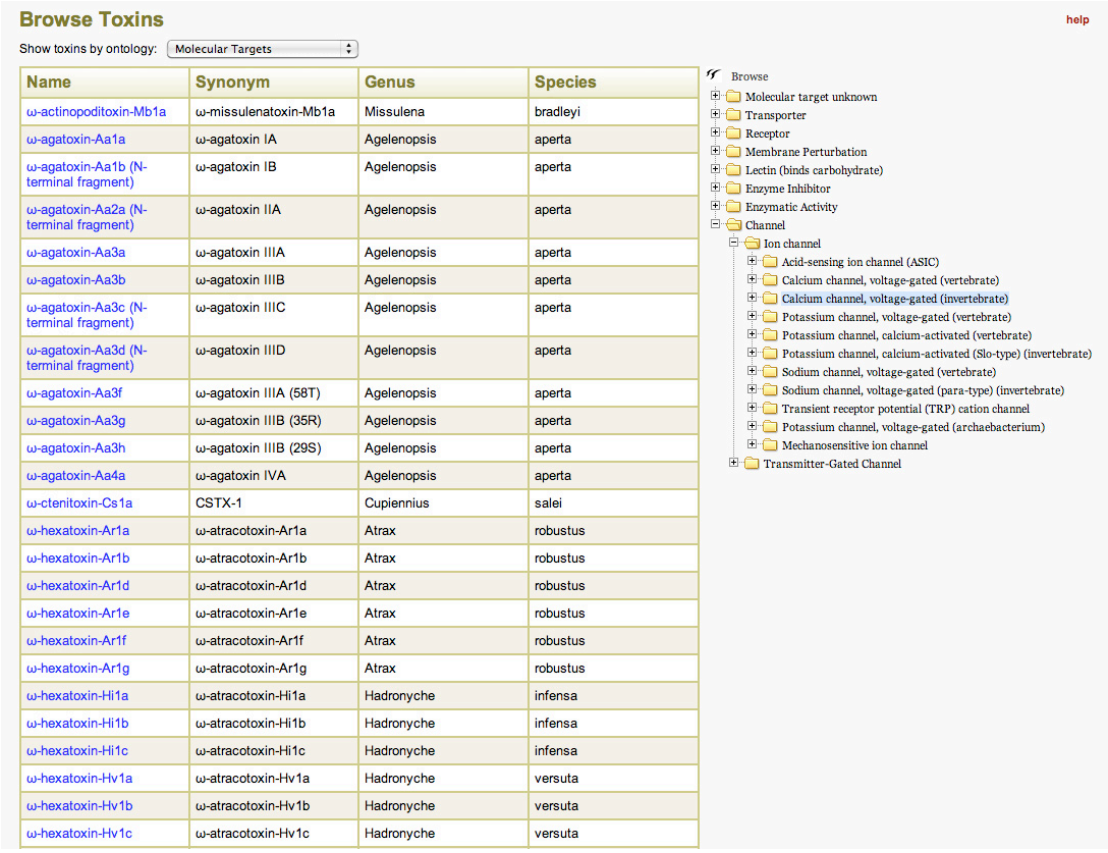


Figure 11. Browsing ArachnoServer using the molecular target ontology to identify all toxins that target invertebrate voltage-gated calcium channels.

Each toxin has its own 'Toxin Card'

Each toxin in the database is provided with its own page, or 'toxin card'. Toxin cards are accessed from the results page of any *Search* or *Browse* (or directly if the ArachnoServer toxin ID is known). The toxin card summarises all available information for this toxin in a single place. It has been designed to minimise information clutter by collapsing categories of information that are likely to be only of interest to particular groups of researchers. As an example, Figure 12 shows the toxin card for κ-theraphotoxin-Gr1a. The following sections explain

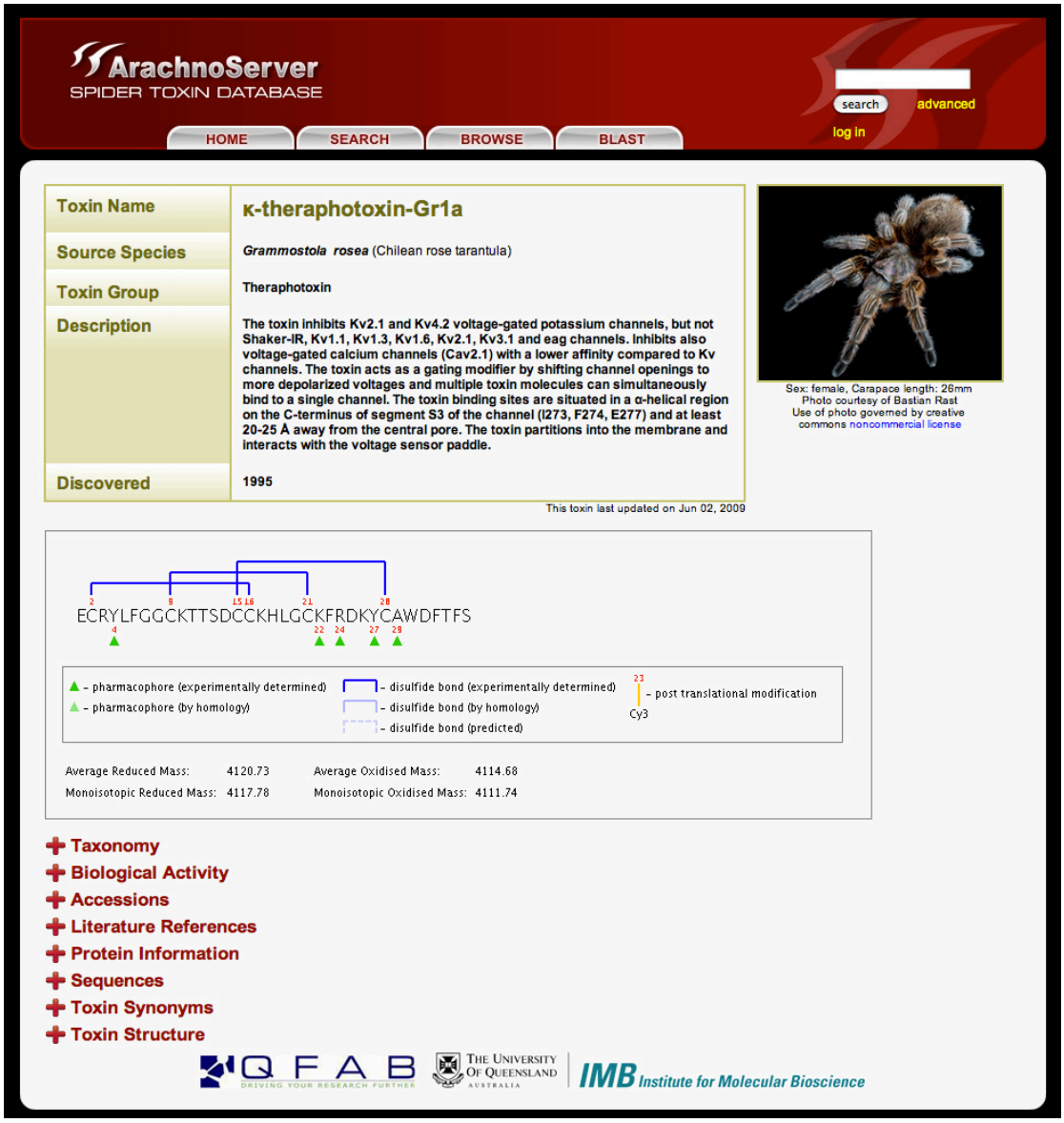


Figure 12. Toxin card for κ-theraphotoxin-Gr1a. All information about this toxin is summarised on this page.

General toxin information

General information about a toxin is displayed at the top of the card; this includes the toxin name, source species (and the common name), the group this toxin belongs to, a general description of the toxin, and the year the toxin was discovered (note that this has been carefully curated from literature information, and may differ from the year the toxin was deposited in a public database).

For most toxins, ArachnoServer also displays a high-resolution image of the source spider. The legend below the photo lists the photographer and provides the prosoma length for scale. Click on the thumbnail image to view a high-resolution version. Images are free for academic use according to the creative commons noncommercial license.

Schematic of the mature toxin

Below the general information for the toxin is a dynamically generated image displaying information about the mature toxin. It shows, where available, the disulfide-bond framework, pharmacophore residues, and posttranslational modifications (Figure 13).

Mousing over the sequence will popup the residue number and amino acid, while mousing over a posttranslational modification will popup the residue number and type of modification (not shown). Note that all residue numbers displayed begin from the start of the mature toxin sequence.

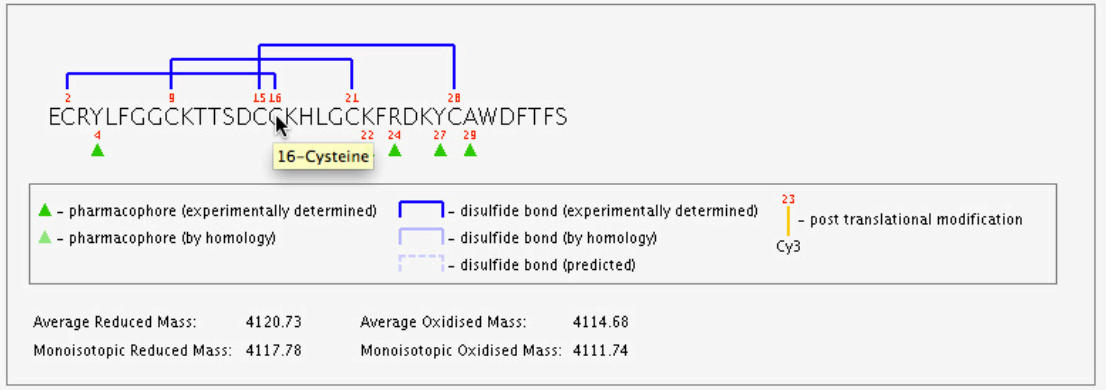


Figure 13. Dynamically generated image of the mature toxin sequence, showing location of disulfide bonds (horizontal blue lines) and pharmacophore residues (green triangles).

Several categories of evidence for disulfide bond annotations are visually displayed. In the figure above, all disulfide bonds have been experimentally determined. However, if this is not the case, they will appear as either defined '*by homology*' or '*predicted*' (see the key below the sequence, where each evidence level is displayed slightly differently in the image). Likewise, two categories of evidence for pharmacophore are displayed, 'experimentally determined' and 'by homology'. The mature toxin displayed above does not have any identified posttranslational modifications.

ArachnoServer also displays information on the mass of the mature toxin sequence within the sequence image. Masses displayed are the average and monoisotopic reduced mass, as well as the average and monoisotopic oxidized masses. This information is also searchable from the advanced search interface.

Toxin card sub-categories

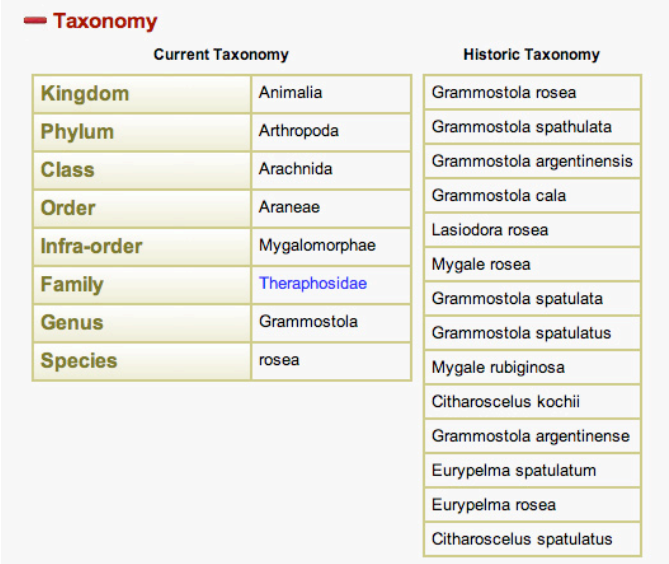
To minimise clutter on the screen and allow for fast access to required information, sub-categories of toxin data have been collapsed within the toxin card (Figure 14).

Figure 14. Sub-categories of information available for a typical toxin. Clicking the category heading will make the information from this category appear.

Each category is expandable by clicking the category heading. If data is not available for a toxin from a particular category, the category will not appear (for example if no solved PDB structure exists, then the '*Toxin Structure*' category will not appear). Following is a description of the information contained within each of the category.

Taxonomy

This section lists the current and historic taxonomy (sourced from the World Spider Catalog). Figure 15 shows an example. As can be seen in this example, *Grammostola rosea* has a long history of taxonomic reclassification. In cases where researchers may not be aware of these changes, or when information from the literature is outdated, the advanced search interface of ArachnoServer enables searching for toxins using this historic taxonomy. The current taxonomic family (Theraphosidae, for this example, in blue) links to the World Spider Catalog entry for that family.



Taxonomy

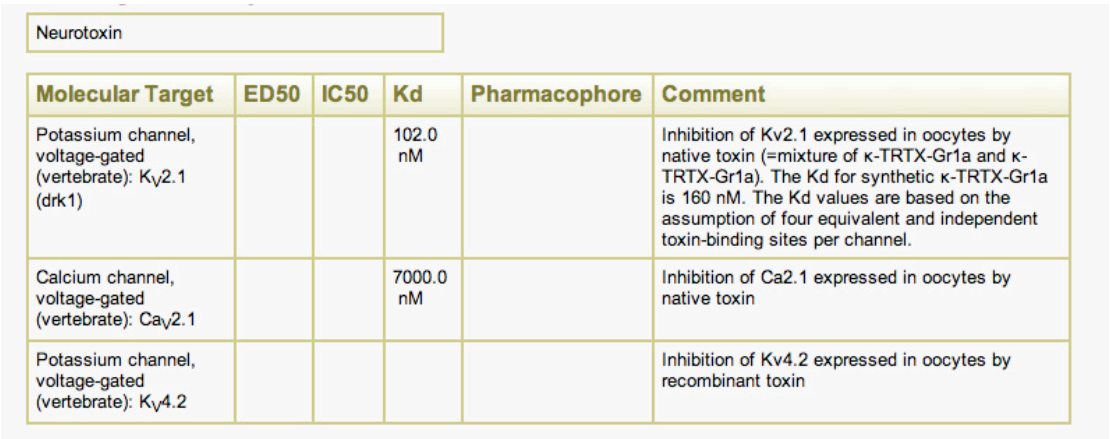
Current Taxonomy	
Kingdom	Animalia
Phylum	Arthropoda
Class	Arachnida
Order	Araneae
Infra-order	Mygalomorphae
Family	Theraphosidae
Genus	Grammostola
Species	rosea

Historic Taxonomy
Grammostola rosea
Grammostola spathulata
Grammostola argentinensis
Grammostola cala
Lasiodora rosea
Mygale rosea
Grammostola spatulata
Grammostola spatulatus
Mygale rubiginosa
Citharoscelus kochii
Grammostola argentinense
Eurypelma spatulatum
Eurypelma rosea
Citharoscelus spatulatus

Figure 15. *Taxonomy Section* of the toxin card, showing the current and historic taxonomy (both sourced from the World Spider Catalog).

Biological activity

This section reports all biological activities (for example neurotoxin, lectin, antimicrobial and others), as well as known molecular targets. Where available and relevant, the ED_{50}, IC_{50}, and K_d values are reported for each molecular target, accompanied with a comment (Figure 16).



Neurotoxin

Molecular Target	ED50	IC50	Kd	Pharmacophore	Comment
Potassium channel, voltage-gated (vertebrate): $K_V2.1$ (drk1)			102.0 nM		Inhibition of Kv2.1 expressed in oocytes by native toxin (=mixture of κ-TRTX-Gr1a and κ-TRTX-Gr1a). The Kd for synthetic κ-TRTX-Gr1a is 160 nM. The Kd values are based on the assumption of four equivalent and independent toxin-binding sites per channel.
Calcium channel, voltage-gated (vertebrate): $Ca_V2.1$			7000.0 nM		Inhibition of Ca2.1 expressed in oocytes by native toxin
Potassium channel, voltage-gated (vertebrate): $K_V4.2$					Inhibition of Kv4.2 expressed in oocytes by recombinant toxin

Figure 16. The *Biological Activity* section of a toxin card displays information about known biological activities and molecular targets.

The list of biological activities in ArachnoServer is under constant curation. Table 2 lists the biological activities currently recognised in ArachnoServer and the number of toxins associated with each activity.

Table 2. List of biological activities recognized in ArachnoServer at the time of writing, with the number of toxins currently associated with that particular activity.

Biological Activity	Toxin Count
Antiarrhythmic	1
Antimicrobial	40
Antinociceptive	2
Antiparasitic	9
Cytolytic	81
Dermonecrotic	46
Hemolytic	49
Kinin-like	1
Lectin	4
Neurotoxin	126
Neurotoxin: Lethal	114
Neurotoxin: Paralytic	125
Presynaptic neurotoxin	4
Protease activity	2
Protease inhibitor	3

Database accessions

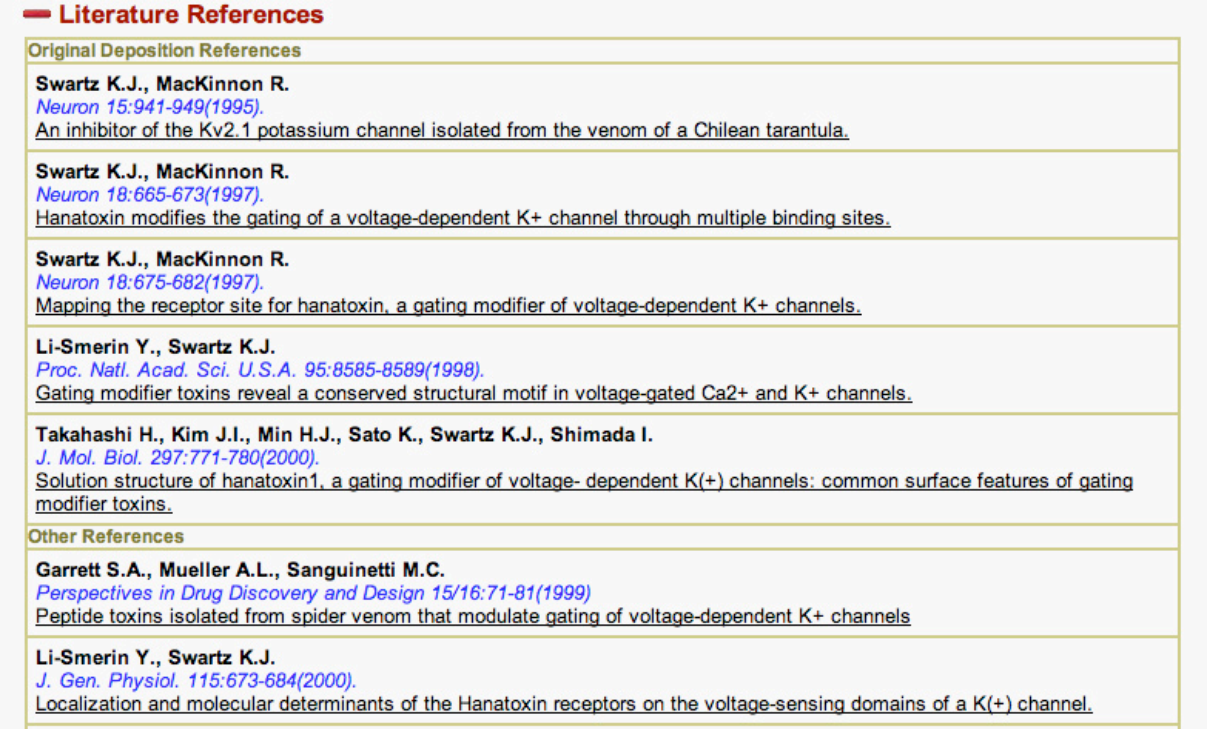
ArachnoServer stores all relevant and current external database cross-references and accession numbers for curated toxins in the database. These external entries can be accessed from the *Accessions* section in each toxin card. Where possible, each of the accessions is hyperlinked to their entry in the external database. Table 3 lists all external databases ArachnoServer links to and the current URL for that database. This list will continue to grow as more toxins are characterised and more databases store toxin-related information.

Table 3. List of all external databases with entries linked to toxins in ArachnoServer.

Database	URL
BioMagResBank	www.mbrb.wisc.edu
BRENDA	www.brenda-enzymes.org
EMBL	www.embl.org
ExPASy - PROSITE	expasy.org/prosite
Gene Ontology (GO)	www.geneontology.org
Gene3D	gene3d.biochem.ucl.ac.uk/Gene3D/
HSSP	swift.cmbi.ru.nl/gv/hssp/
MEROPS	merops.sanger.ac.uk
PANTHER	www.pantherdb.org
Pfam	pfam.sanger.ac.uk
PRINTS	www.bioinf.manchester.ac.uk/dbbrowser/PRINTS
ProDom	prodom.prabi.fr
Protein Data Bank (PDB)	www.pdb.org
SMART	smart.embl-heidelberg.de
UniProt Knowledge Base	www.uniprot.org

Literature references

Every toxin in ArachnoServer has been carefully curated using a wide variety of information. Much of the additional information within the database is sourced from the reading and recording of results from journal articles and patents. When this information is incorporated into the database, the curator also records the citation for that article or information about the patent.



Literature References

Original Deposition References
Swartz K.J., MacKinnon R. *Neuron 15:941-949(1995).* An inhibitor of the Kv2.1 potassium channel isolated from the venom of a Chilean tarantula.
Swartz K.J., MacKinnon R. *Neuron 18:665-673(1997).* Hanatoxin modifies the gating of a voltage-dependent K+ channel through multiple binding sites.
Swartz K.J., MacKinnon R. *Neuron 18:675-682(1997).* Mapping the receptor site for hanatoxin, a gating modifier of voltage-dependent K+ channels.
Li-Smerin Y., Swartz K.J. *Proc. Natl. Acad. Sci. U.S.A. 95:8585-8589(1998).* Gating modifier toxins reveal a conserved structural motif in voltage-gated Ca2+ and K+ channels.
Takahashi H., Kim J.I., Min H.J., Sato K., Swartz K.J., Shimada I. *J. Mol. Biol. 297:771-780(2000).* Solution structure of hanatoxin1, a gating modifier of voltage- dependent K(+) channels: common surface features of gating modifier toxins.
Other References
Garrett S.A., Mueller A.L., Sanguinetti M.C. *Perspectives in Drug Discovery and Design 15/16:71-81(1999)* Peptide toxins isolated from spider venom that modulate gating of voltage-dependent K+ channels
Li-Smerin Y., Swartz K.J. *J. Gen. Physiol. 115:673-684(2000).* Localization and molecular determinants of the Hanatoxin receptors on the voltage-sensing domains of a K(+) channel.

Figure 17. Example literature table from the toxin card for κ-theraphotoxin-Gr1a.
Each citation is linked to its Pubmed record (where available).

Literature references include research articles, details about sequence deposition dates, and patent information. The references are normally divided into two groups, those for the original deposition, and additional references. Where possible, each of the citations will link to the PubMed page for that article.

ArachnoServer currently contains over 1650 citations, sourced from more than 400 distinct journal articles in 80 different journals (excluding patent records and references to original deposition details in public databases). All of these literature articles, including the names of original deposition authors, as well as patent information, are directly searchable from the advanced search interface.

Protein information

This section contains detailed information on the primary structure of each toxin, such as schematics of the signal, propeptide, and mature toxin sequences (if known), the disulfide-bond scaffold, and any posttranslational modifications (Figure 18).

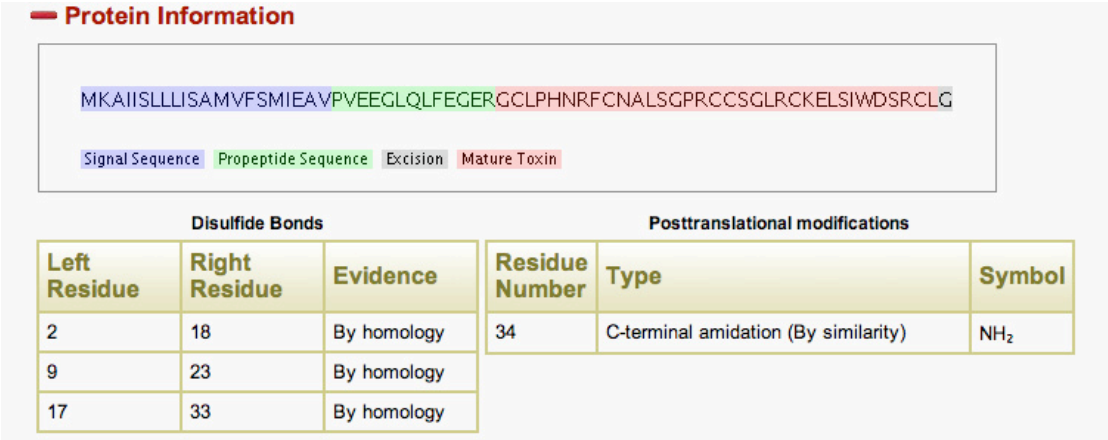


Figure 18. Protein information for the toxin U2-agatoxin-Ao1g from *Agelena orientalis* showing the signal, propeptide, and mature toxin sequences, as well the C-terminal Gly residue that is excised in the process of C-terminal amidation. The protein information section also lists locations and available evidence for disulfide bonds and posttranslational modifications.

In some spiders, different prepropeptide transcripts can encode the same mature toxin (i.e., the transcripts have the same mature toxin sequence but have at least one difference in the signal or propeptide region). As each toxin in ArachnoServer is uniquely defined by its mature toxin sequence, all such transcripts will be stored on the same toxin card. The *Protein Information* tab provides a quick schematic overview of the range of transcripts encoding the same mature toxin. Figure 19 illustrates one such example.

Figure 19. Protein information for the toxin κ-hexatoxin-Hv1c (*Hadronyche versuta*) showing multiple transcripts with slightly different signal and propeptide sequences.

Sequences

All available sequences for a given toxin will be displayed in this section, such as protein sequences (either with or without signal and propeptide sequences), as well as mRNA and DNA sequences if known (Figure 20). Note that, at the time of writing, there is very little information available on the structure of spider-toxin genes. ArachnoServer currently holds 926 protein sequences, 632 mRNA sequences and 29 DNA sequences.

The sequences section is a convenient location to perform similarity searches of ArachnoServer for related toxins. More information on this is provided in the section below on '*Running Similarity Searches*'.

Sequences

Peptide Sequences		
>as:κ-hexatoxin-Hv1c_1|sp:P82228 Insecticidal toxin from Blue Mountains funnel-web spider Hadronyche versuta AICTGADRPCAACCPCCPGTSCKAESNGVSYCRKDEP	Full BLAST	BLAST mature toxin only
>as:κ-hexatoxin-Hv1c_2 Translation κ-hexatoxin-Hv1c insecticidal toxin (XenFW386) from Blue Mountains funnel-web spider Hadronyche versuta. MNTATCFIVLLDVATVIGGIEAGESDMRKDVMGLFRRAICTGADRPCAACCPCCPGTSCK AESNGVSYCRKDEP	Full BLAST	BLAST mature toxin only
>as:κ-hexatoxin-Hv1c_3 Translation κ-hexatoxin-Hv1c insecticidal toxin (XenFW393) from Blue Mountains funnel-web spider Hadronyche versuta. MNTATCFIVLLVVATVIGGIEAGEFDMRKDVMGLFRRAICTGADRPCAACCPCCPGTSCK AESNGVSYCRKDEP	Full BLAST	BLAST mature toxin only
>as:κ-hexatoxin-Hv1c_4 Translation κ-hexatoxin-Hv1c insecticidal toxin (XenFW402) from Blue Mountains funnel-web spider Hadronyche versuta. MNTATCFIVFLVVATVIGGIEAGESDMRKDVMGLFRRAICTGADRPCAACCPCCPGTSCK AESNGVSYCRKDEP	Full BLAST	BLAST mature toxin only
mRNA Sequences		
>as:κ-hexatoxin-Hv1c_2 cDNA sequence (XenFW386) for insecticidal toxin κ-hexatoxin-Hv1c from the spider Hadronyche versuta. ATGAATACTGCTACATGTTTCATCGTTCTTTTGGATGTGGCGACTGTCATCGGAGGCATT GAAGCAGGAGAATCTGATATGAGAAAAGATGTCATGGGATTATTTCGCCGAGCTATTTGC ACTGGAGCCGACAGACCGTGCGCGGCGTGCTGCCCGTGCTGCCCAGGGACCTCGTGCAAA GCAGAATCAAACGGTGTTTCTTATTGCAGGAAAGACGAACCTTGAATCCCAACTCAAATC ATTTCCCTTCGCATGATATTTTCTATTAATGTCTTTTATCGTTGCAAGGAAATGTCAATG GTTATGTTGAACGTTATGTAAAACGTTGACTATGAAATAAAATGTGAACAATAATTAAAT ACCTGAAAAAAAAAAAAAAAAAAAAAA	Full BLAST	
>as:κ-hexatoxin-Hv1c_3 cDNA sequence (XenFW393) for insecticidal toxin κ-hexatoxin-Hv1c from the spider Hadronyche versuta. ATGAATACTGCTACATGTTTCATCGTTCTTTTGGTTGTGGCGACTGTCATCGGAGGCATT GAAGCAGGAGAATTTGATATGAGAAAAGATGTCATGGGATTATTTCGCCGAGCTATTTGC ACTGGAGCCGACAGACCGTGCGCGGCGTGCTGCCCGTGCTGCCCAGGGACCTCGTGCAAA GCAGAATCAAACGGTGTTTCTTATTGCAGGAAAGACGAACCTTGAATCCCAACTCAAATC ATTTCCCTTCGCATGATATTTTCTATTAATGTCTTTTATCGTTGCAAGGAAATGTCAATG GTTATGTTGAACGTTATGTAAAACGTTGACTATGAAATAAAATGTGAACAATAATTAAAT CCCTTAAAAAAAAAAAAAAAAAAAAAC	Full BLAST	

Figure 20. Sequences stored in ArachnoServer for the toxin κ-hexatoxin-Hv1c (some mRNA sequences have been excluded for brevity). All sequences can be directly sent to the BLAST page from here. For protein sequences, either the full sequence, or just the mature toxin sequence, can be used as the search query.

Toxin synonyms

This section displays all known synonyms for the toxin, as well as the current recommended names and abbreviations. For example, the toxin κ-hexatoxin-Hv1c has the following synonyms in ArachnoServer:

Toxin Synonyms

Synonym	Type
κ-hexatoxin-Hv1c	Recommended full name
κ-HXTX-Hv1c	Recommended abbreviation
κ-atracotoxin-Hv1c	Synonym
Janus-faced-atracotoxin-Hv1c	Synonym
κ-ACTX-Hv1c	Synonym (abbreviation)
J-atracotoxin-Hv1c	Synonym (abbreviation)
J-ACTX-Hv1c	Synonym (abbreviation)

Figure 21. A list of synonyms for κ-hexatoxin-Hv1c.

Recommended abbreviations are created using a four or five letter abbreviation of the generic toxin name as described at http://www.venomics.org/nomenclature/. All synonyms are searchable from the advanced search interface.

Toxin structure

ArachnoServer also stores information on solved toxin structures that have been submitted to the Protein Data Bank (PDB). Currently, 43 structures are available, and these can be quickly identified using the advanced search field '*number of solved PDB structures*'. To assist in the viewing of structures, ArachnoServer has an embedded molecular structure viewer (Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/). In order to use the viewer, your browser must support java applets (as do all modern web browsers). Figure 22 shows one of the solved structures of the toxin μ-theraphotoxin-Hhn1b from the Chinese Black Earth Tiger tarantula *Haplopelma hainanum* . To view an alternative structure, simply select the appropriate button at the top of the viewer; these buttons display the PDB accession number for the structure.

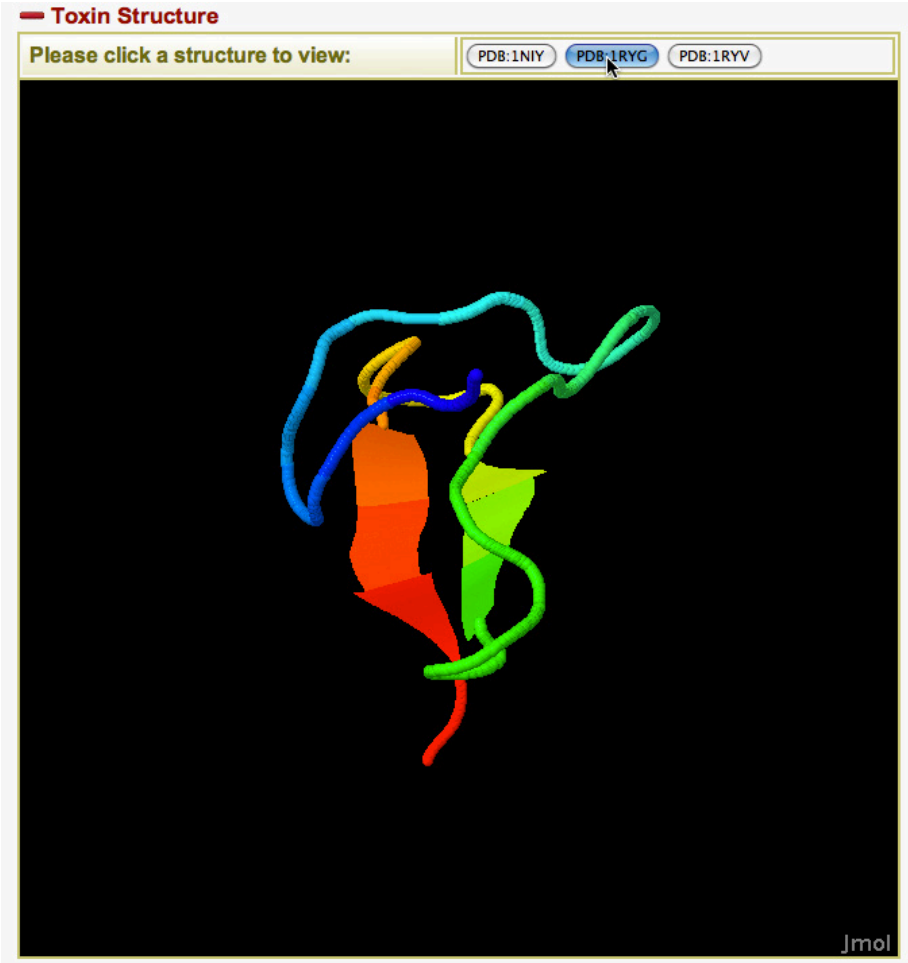


Figure 22. Structure (PDB:1RYG) of the toxin μ-theraphotoxin-Hhn1b, viewed in ArachnoServer using the embedded molecular structure viewer (Jmol). Multiple structures have been solved for this toxin; each of these can be viewed by clicking the appropriate button displaying the PDB accession number at the top of the structure viewer.

Many alternative viewing options are available in Jmol (so long as you have a three-button mouse). Right click the viewing panel to open a menu, then select the view and style you prefer. Alternative surfaces can also be added. For a full description of what Jmol can offer, please visit their web site (http://www.jmol.org).

Running similarity searches 

ArachnoServer makes available all toxin sequences for similarity searches using BLAST (Basic Local Alignment Search Tool). Click the BLAST tab, and then paste into the text area your sequence of interest. Select the type of query sequence (*Protein* or *DNA*) and the database to search against (either *Peptide* or *Nucleic Sequences*). Once both of these selections have been made, you can select the BLAST program to use. For a description of BLAST programs, see here. All main BLAST programs are available in ArachnoServer. Figure 23 shows the BLAST interface and some of the available options.

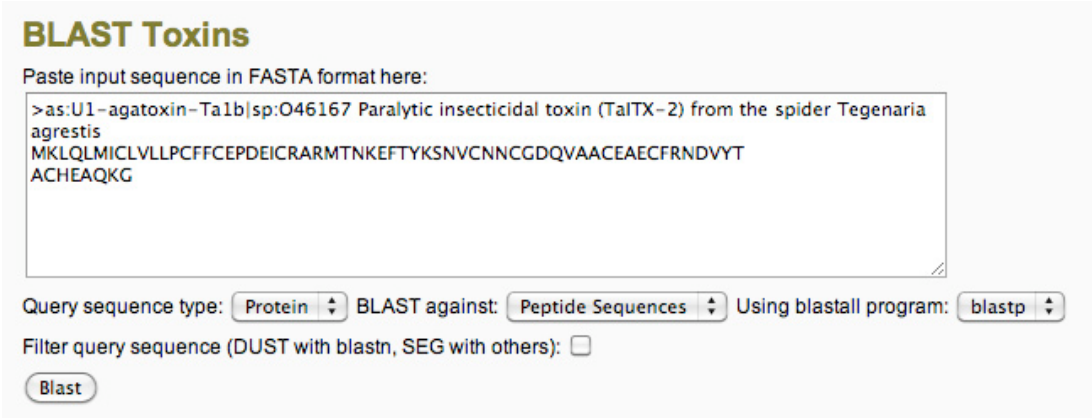


Figure 23. BLAST form in ArachnoServer populated with a toxin from the database. Note that SEG and DUST filtering is off by default.

Alternatively, under the *Sequences* section of each toxin card, each sequence can be sent to the BLAST search form and run using a single click. This is a convenient way of quickly identifying toxins similar to those already in the database. For protein sequences, both the full sequence, including the signal and propeptide sequence where available, or just the mature toxin sequence can be selected as the BLAST query. As toxins are uniquely identified by their mature toxin sequence, this second option provides a mechanism to enrich the alignment results for mature toxins only.

Due to the often very short length of mature toxin sequences, when running BLAST on toxin data, we recommend you leave the low compositional complexity filter off (DUST for BLASTn, and SEG for other BLAST programs).

The BLAST results will be formatted in HTML and contain links to ArachnoServer toxin cards for those toxins that are most similar to the query sequence. If those toxins also have links to UniProtKB/Swiss-Prot, these will also be available in the BLAST result (Figure 24).

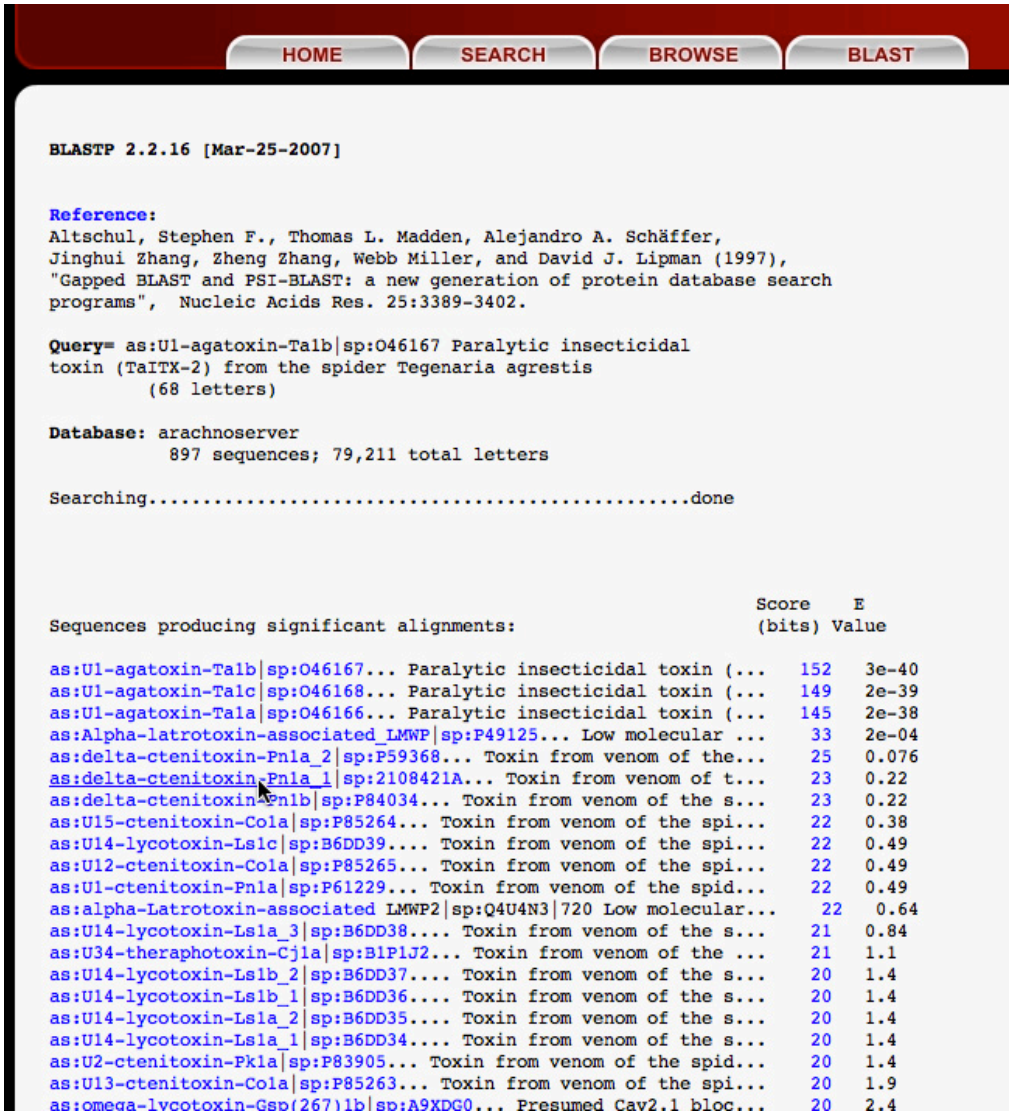


Figure 24. BLAST results as formatted in ArachnoServer. Each hit is linked to the corresponding toxin card in the database, as well as to Uni-ProtKB/Swiss-Prot records where available.

The links in the score column provide a quick mechanism to access a text representation of the alignment for a particular hit.

BLAST databases are stored on the server, and are updated every 15 minutes to ensure that the similarity searches most accurately represent the curated data in ArachnoServer.

Linking to ArachnoServer

ArachnoServer entries are externally identifiable by accessions with the letters 'AS' followed by six digits. These accessions can be used to link to ArachnoServer; for example, the following hypertext reference links to the toxin ω-ctenitoxin-Cs1a from the wandering spider *Cupiennius salei:*

http://www.arachnoserver.org/toxincard.html?id=AS000292

The non-zero digits are retrieved from the toxin ID, found in the URL for all toxins, then the ID is padded with zeros in front to make a six digit number. Toxin IDs in ArachnoServer are persistent and will never change for that toxin.

UniProtKB/Swiss-Prot mapping to ArachnoServer

Data from ArachnoServer is now being used to assist the curators at UniProtKB/Swiss-Prot to provide the most up-to-date descriptions for spider toxin characterisation. The rational toxin names recommended in ArachnoServer are also being adopted by UniProtKB/Swiss-Prot, and are directly searchable from their interface.

You can map UniProtKB/Swiss-Prot identifiers to ArachnoServer toxin entries, or vice-versa, by using the UniProtKB/Swiss-Prot ID Mapping Service; just select the 'ArachnoServer' database under the 'Organism Specific Databases' section.