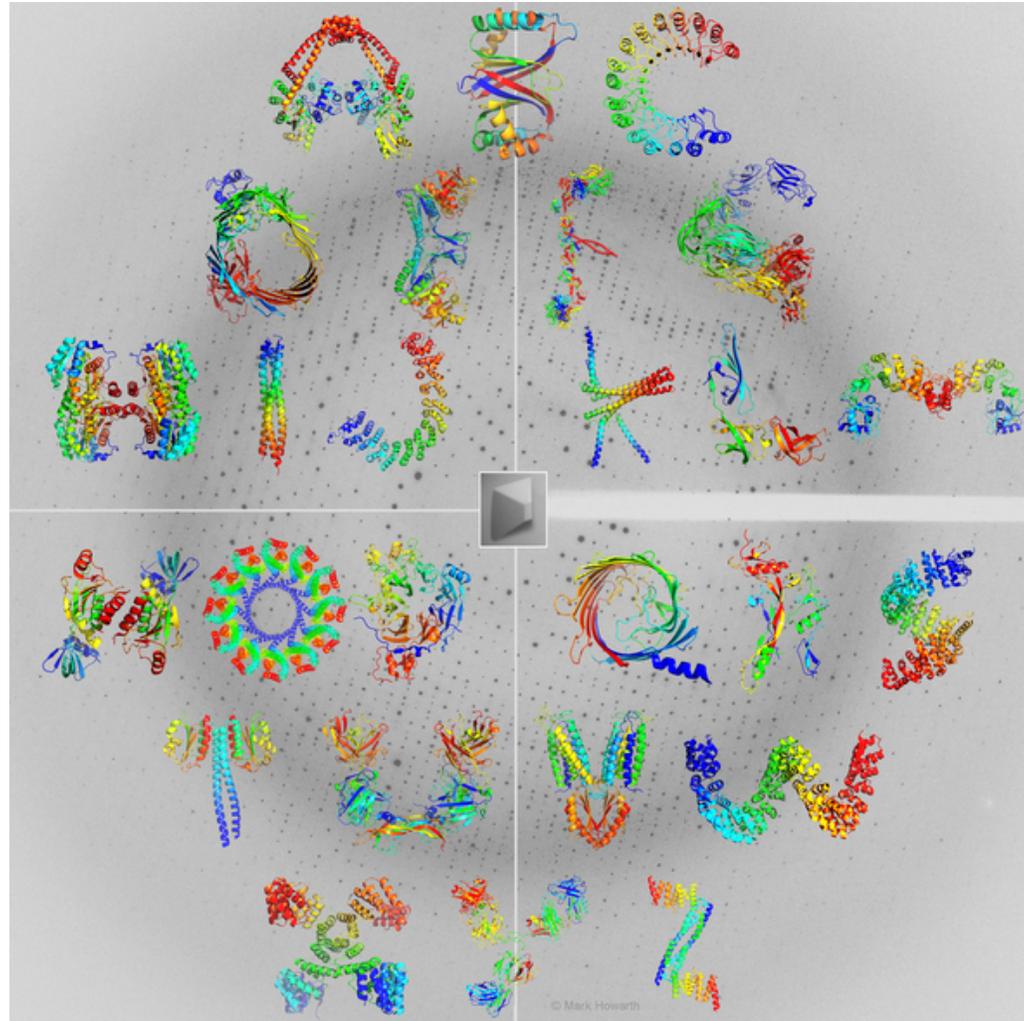


Databases and online tools for protein structure

Shaikh Nausad Hossain

Proteins



Why Protein structure?



- Understand the function
- Protein ligand interaction for drug discovery

Protein: Dihydrofolate reductase
Drug: Trimethoprim

Protein Data Bank (PDB)

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB An Information Portal to 120057 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands

Advanced Search | Browse by Annotations | Search History (2) | Previous Results (24)

PDBe-101 PDBe EMDatabank StructuralBiology Worldwide Protein Data Bank Foundation

Welcome

- Deposit
- Search
- Visualize
- Analyze
- Download
- Learn

A Structural View of Biology

This resource is powered by the Protein Data Bank archive—information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

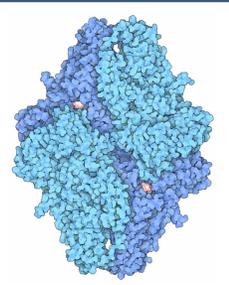
The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Video Challenge Awards

More Info

- 1st Place: West Windsor Plainsboro South HS - "Sweet Signals"
- Viewer's Choice: McNair Academic HS - "It's Me, You, and Glucose"

June Molecule of the Month



Beta-galactosidase

- ❑ 120057 structures deposited
- ❑ 111438 protein structures
- ❑ 2518 Membrane proteins

Latest Entries

As of Tuesday, Jun 28



5D90

Crystal structure of HiNmIR, a MerR family regulator lacking the sensor domain, bound to promoter DNA

New Features

February 2016 Release

- Redesigned Search Results Page**
New Organization. Improved Layout. Clean. Usable. Faster Load.
- Electron Density Map Visualization
- Genetic Variation Track on Protein Feature View
- Wild Type Search

October 2015 Release

- Redesigned Structure Summary Page**
New Organization. Improved Layout. Clean. Usable. Simple.

News

Publications

- 2016 ISMB Conference**
Join us for posters, presentations, hackathons, and honors at the Intelligent Systems for Molecular Biology (ISMB) Conference (July 8-12). - 06/28/16
- Access to wwPDB and RCSB PDB websites - 06/21/16
- Award-Winning Videos about Structural Biology and Diabetes - 06/14/16
- AutoDep Deposition System to Be Retired, Effective August 25th 2016**
- ADIT-NMR and EMDep Deposition Systems to Be Retired, Effective September 30th 2016**

Protein sequence

- Prediction of domains
- Secondary structure prediction

Online tools for protein

ProtParam tool

ProtParam ([References](#) / [Documentation](#)) is a tool which allows the computation of various physical and chemical parameters for a given protein stored in [Swiss-Prot](#) or [TrEMBL](#) or for a user entered protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) ([Disclaimer](#)).

Please note that you may only fill out **one** of the following fields at a time.

Enter a Swiss-Prot/TrEMBL accession number (AC) (for example **P05130**) or a sequence identifier (ID) (for example **KPC1_DROME**):

Or you can paste your own amino acid sequence (in one-letter code) in the box below:

- Molecular weight
- Isoelectric point
- Amino acid composition
- Atomic composition
- Extinction coefficient

- Half life
- Instability index
- Hydrophobicity

Domain prediction



ScanProsite Results Viewer

Output format: Graphical view - this view shows ScanProsite results together with ProRule-based predicted intra-domain features [\[help\]](#).

Hits for all PROSITE (release 20.127) motifs on sequence USERSEQ1 :

found: 4 hits in 1 sequence

USERSEQ1 (608 aa)

```
MKWLKQLQSLHTKLVIVYVLLIIIGMQIIIGLYFTNNLEKELLDNFKKNITQYAKQLEISIEKVYDE
KGSVNAQKDIQNLLSEYANRQEIGEIRFIDKQII IATTKQSNRSLINQKANDSSVQKALSLGQSN
DHLILKDYGGGKDRVWVYNI PVKVDKRVIGNIYIESKINDVYNQLNINQIFIVGTAISLLITVIL
GFFIARTITKPI TDMRNQTVEMSKGNYTQRVKI YGNDEIGELALAFNNLSKRVQEAQANTESEKCR
LDSVITHMSDGI IATDRRGRIRIVNDMALKMLGMAKEDIIGYYMLSVLSLEDEFKLEEIQENNSF
LLDLNEEGLIARVNFSTIVQETGFVTGYIAVLHDVTEQQQVERERREFVANVSHELRTPLTSMNS
YIEALEEGAWKDEELAPQFLSVTREETERMIRLVNDLLQLSKMDNESDQINKEIIDFNMFINKIIN
RHEMSTKDTTFIRDIPKKTIFTEFDPDKMTQVFDNVI TNAMKYSRGDKRVEFHVQKPNLYNRMTIR
IKDNGIGIPINKVDKIFDRFYRVDKARTRKMGGTGLGLAISKEIVEAHNGRIWANSVEGQGTSIFI
TLPCEVIEDGDWDE
```

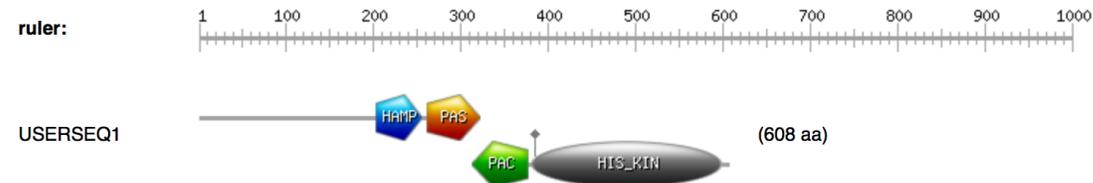
Legend:



Please note that the graphical representations of domains displayed hereafter are for illustrative purposes only, and that their colors and shapes are not intended to indicate homology or shared function. For more information about how these graphical representations are constructed, go to <http://prosite.expasy.org/mydomains/>.

hits by profiles: [4 hits (by 4 distinct profiles) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.



Secondary structure prediction



CFSSP: Chou & Fasman Secondary Structure Prediction Server

Home Blog Tools Academic Contact Mail

This server predicts secondary structure of protein from the amino acid sequence. In this server, Chou & Fasman algorithm has been implemented.

— Enter the protein sequence (in fasta format) —

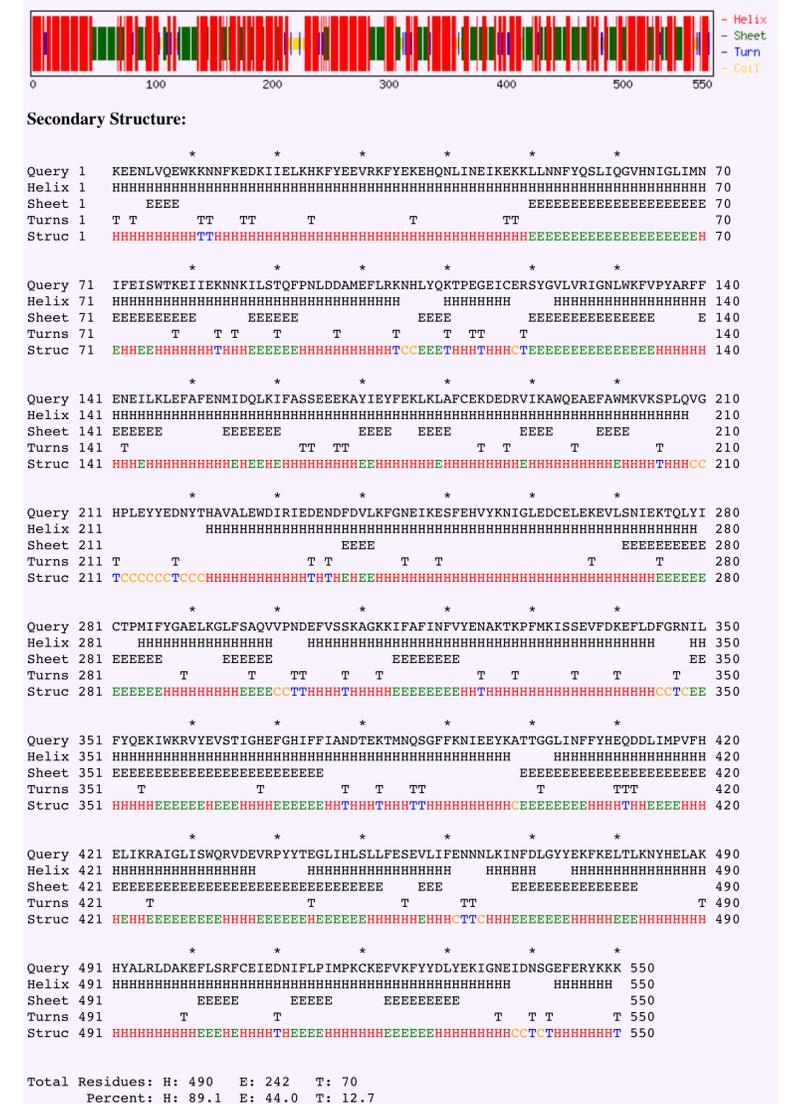
```
> MNNFKEIAKLVRYKERNNALYEFLDKEDVGEYFRSLISLSELKQDKTTMLAILRRLIDL
KEENLVQEWKKNNFKEDIIEKHKFYEEVRKFYEKEHQNLNIEKPKLLNNFYQSLIQ
GVHNI GLIMNIFEISWTKEIEKNNKILSTQFPNDDAMEFLRKNHLYQKTPEGEICERS
YGVLVIRIGNLWKFVYPYARFFENEILKLEF AFENMIDQLKIFASSEEEKAYIEYFEKLLA
FCEKDEDRVIKAWQEAFAWMKVKSPLQVGHPLYEDNYTHAVALEWDIRIEDENDFDV
LKFGNEIKESFEHVYKNIGLEDCELEKEVLSNIEKTQLYICTPMIFYGAELKGLFSAQVV
PNDEFVSSKAGKKIFAFINFYENAKTKPFMKISSEVDFKELDFGRNILFYQEKIWKRV
MELVETIQHEEQLWEEFLNDEKFMAGGEEFKWEEFYATQQLMEEYHEQDRLMDEFL
```

CLEAR

PREDICT

Reference:

1. Peter Y. Chou, and Gerald D. Fasman. Prediction of protein conformation. *Biochemistry*. 1974 Jan; **13**(2), pp 222-245.
2. Peter Y. Chou, and Gerald D. Fasman. Conformational parameters for amino acids in helical, β -sheet, and random coil regions calculated from proteins. *Biochemistry*. 1974 Jan; **13**(2): pp 211-222.



Functional analysis from structure

EMBL-EBI Services Research Training About us

ProFunc results for 4mn5

Services > Structure Databases > ProFunc > Results

ProFunc



ProFunc id: 4mn5

Header details

Structure: CRYSTAL STRUCTURE OF PAS DOMAIN OF S. AUREUS YYCG

Source: Saphylococcus aureus. Organism_taid: 1280, Genevick, walk, yycg, Escherichia coli. Expression_system_taxid: 511693, Ppge2tt

Date: 10 Sep 13

Author(s): N. Shaikh, R. Hvorup, B. Winnen, B. M. Collins, G. F. King

Resolution: 2.000Å

Waters: 76

Jmol RasMol

Summary of predicted function

Protein name terms

kinase (36.53) os=staphylococcus (30.31) sensor (27.30) histidine kinase (19.51) aureus (14.67) domain (13.04) pas (10.94) kinase walk (10.84)

Gene Ontology (GO) terms

Cellular component: (0)

Biological process: regulation of transcription (0.80) DNA-dependent (0.80) biological regulation (0.80) regulation of biological process (0.80) regulation of cellular process (0.80)

Biochemical function: two-component sensor activity (0.80) catalytic activity (0.80) transferase activity (0.80) transferase activity, transferring phosphorus-containing groups (0.80)

The protein names and GO terms above are the most common terms found in the hits obtained from the analyses below. Each term's score (based on the number of times it occurs independently) is given in brackets. Click on the plus icons for a complete breakdown of which programs, and further, which hits the terms came from.

ProFunc results

Sequence motifs

InterPro scan for sequence motifs. Chains A, B

6 motifs matched in scan against PROSITE, PRINTS, Pfam-A, TIGRFAM, PROFILES and PRODOM motifs

Type	Motif	Name
1. SMART	SM00091	
2. Pfam	PF00989	PAS fold
3. SUPERFAMILY	SSF5785	
4. ProSiteProfiles	PSS0113	PAC domain profile.
5. ProSiteProfiles	PSS0112	PAS repeat profile.

... plus others

Matches to existing PDB structures

PDB Sequence search vs existing PDB entries. Chains A, B

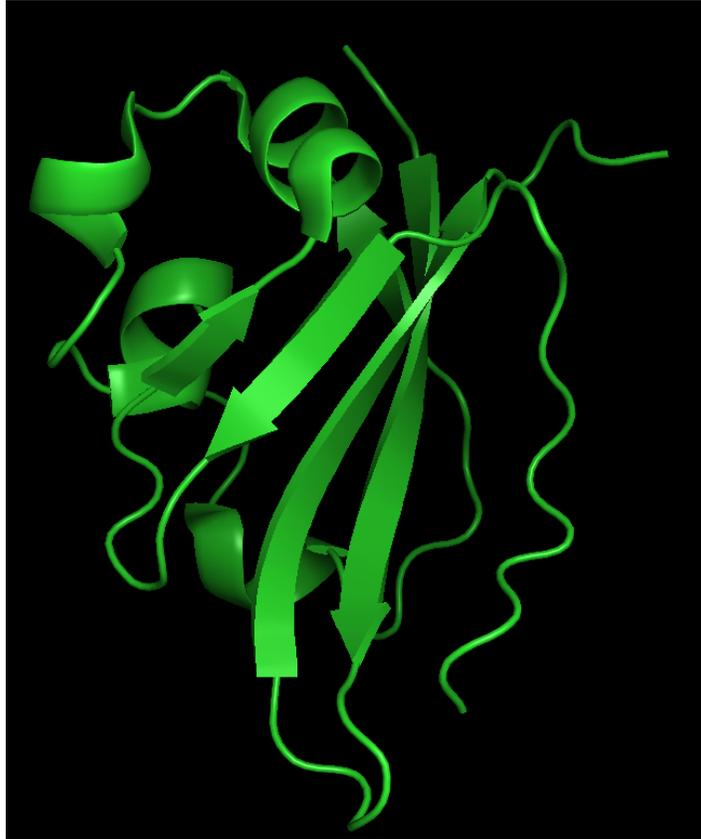
8 matching sequences found by FASTA search

PDB code	E-value	%-tag	id	Overlap	Name
1. 4Is6(A)	1.4e-11	36.111	108	Structure and function of sensor histidine kinase	
2. 3a0r(A)	0.11	24.762	105	Crystal structure of histidine kinase thka (tm1359) in compl response regulator protein trs (tm1360)	
3. 3a0v(A)	1.1	23.958	96	Pas domain of histidine kinase thka (tm1359) (semet, f486mVf489m)	
4. 3a0s(A)	2.1	23.958	96	Pas domain of histidine kinase thka (tm1359)	
5. 3ewk(A)	2.5	34.375	64	Structure of the redox sensor domain of methylococcus capsul (bath) mmos	

... plus others

Input: PDB code

Structure matching



4MN5



4I5S

Sequence search



BLAST search vs Uniprot. Chains A, B

50 matching sequences found by BLAST search

Ref. no.	E-value	%-tag	id	Overlap	Name
1. Q1ZZY3_STAAU	2e-55	96.262	107		Q1ZZY3 VicK (Fragment) OS=Staphylococcus aureus GN=vicK PE=4
2. WALK_STAAE	3e-51	95.327	107		A6QD58 Sensor protein kinase Walk OS=Staphylococcus aureus (strain
3. A0A0E1ACCS_STAAU	4e-51	96.226	106		A0A0E1ACCS Two-component sensor kinase Walk OS=Staphylococcus
4. A0A133S0T2_9FIRM	7e-50	85.047	107		A0A133S0T2 PAS domain S-box protein (Fragment) OS=Veillonella
5. A0A0W3BNW6_STAHA	1e-48	87.850	107		A0A0W3BNW6 PAS domain-containing sensor histidine kinase OS=Staphylococcus

... plus others

Residue conservation analysis



Conservation mapped onto 3D structure.

Matching folds detected by PDBeFold



Matching folds. Chains A, B

24296 significant structural matches

Q-score	Rmsd	No. SSE	Z-score	PDB	Name
1. 0.860	0.74Å	9	13.5	4mn6	Crystal structure of truncated pas domain from s. Aureus yyc
2. 0.432	1.77Å	6	8.3	1g28	Structure of a flavin-binding domain, lov2, from the chimeric phytochrome/phototropin photoreceptor phy3
3. 0.428	2.10Å	5	8.1	1odv	Photoactive yellow protein 1-25 deletion mutant
4. 0.424	2.17Å	6	7.5	5iu1	N-terminal pas domain homodimer of ppar map3k from physcomi patens.
5. 0.422	1.99Å	5	7.7	4hqa	Crystal structure of pas domain from the human erg (herg) po channel

... plus others

Nest analysis



Nest analysis. Chains A, B

4 nests located in the structure

Score	resid.	No. access-	Ave. cleft	Ave. conser-	Residues
1. 1.964	3	0.00	3	0.964	Arg282(A), Gly283(A), Arg284(A)
2. 1.634	3	0.00	4	0.634	Leu296(A), Gly297(A), Mse298(A)
3. 1.000	3	0.00	-	1.000	Thr353(A), Gly354(A), Phe355(A)
4. 0.905	4	0.00	-	0.905	Leu312(A), Ser313(A), Leu314(A), Glu315(A)

Cleft analysis



Surface clefts.

3D functional template searches



Enzyme active site templates.

No hits obtained from any of the 584 enzyme active site templates.



Ligand-binding templates.

1 significant hit out of 94055 ligand-binding templates.

Score	Template	PDB Name
1. 174.117	PXGb0001 1b9l	Crystal structure of 3-amino-5-hydroxybenzoic acid (ahba) by Het Group PXG



DNA-binding templates.

No hits obtained from any of the 5320 DNA-binding templates.

Nest analysis results for 4mn5 (chain A)

Services > Structure Databases > ProFunc > Results

ProFunc

Nests are structural motifs that are often found in functionally important regions of protein structures.

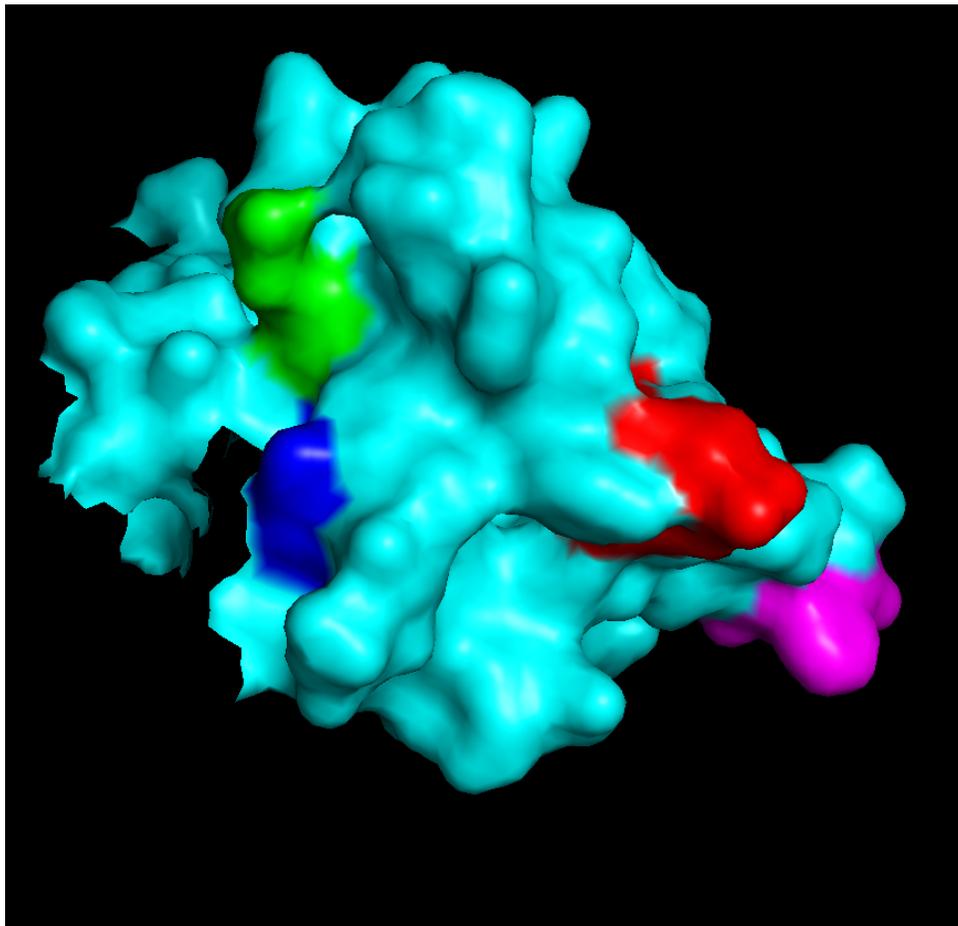
4 nests were located in this chain, as shown below.

Nest	Score	Residue range	View in RasMol	Residue	Ramachandran region	Solvent accessibility	Cleft	Depth in cleft	Residue conservation
1.	1.96	Arg282(A)-Arg284(A)		Arg282(A)	RIGHT	0.00%	3	-	0.89
				Gly283(A)	LEFT	0.00%	-	-	1.00
				Arg284(A)	-	0.00%	-	-	1.00
2.	1.63	Leu296(A)-Mse298(A)		Leu296(A)	RIGHT	0.00%	-	-	0.90
				Gly297(A)	LEFT	0.00%	4	4.96	1.00
				Mse298(A)	-	0.00%	-	-	0.00
3.	1.00	Thr353(A)-Phe355(A)		Thr353(A)	RIGHT	0.00%	-	-	1.00
				Gly354(A)	LEFT	0.00%	-	-	1.00
				Phe355(A)	-	0.00%	-	-	1.00
4.	0.90	Leu312(A)-Glu315(A)		Leu312(A)	RIGHT	0.00%	-	-	1.00
				Ser313(A)	LEFT	0.00%	-	-	0.62
				Leu314(A)	RIGHT	0.00%	-	-	1.00
				Glu315(A)	RIGHT	0.00%	-	-	1.00



Key to table entries

Nest analysis



Possible ligand binding sites

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 toxinologists, can readily access key data relevant to their discipline without being overwhelmed by extraneous information.

All spider toxin entries are sourced from UniProtKB/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., 2008).

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

Toxins (total)	1525
Toxins (curated)	1407
Species	97

Please visit the [download](#) page to download data from ArachnoServer.

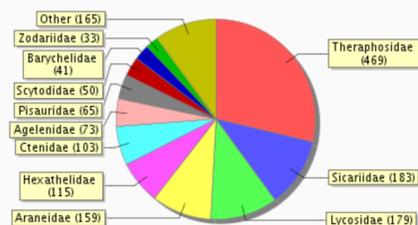
UPDATE: Please try out [Tox|Note](#), our new pipeline to significantly fast track the analysis of spider venom-gland transcriptomes, by isolating all toxin/ toxin-like sequences generated by large Next Generation (NG) sequencing projects.

Tested with:

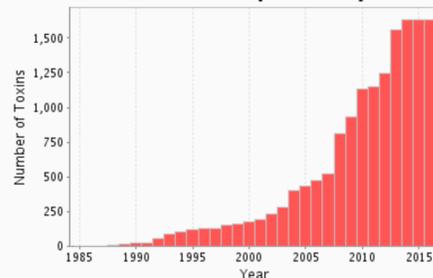


[But please enable javascript](#)

Toxin Distribution by Family



Cumulative Toxins Deposited by Year



If you used ArachnoServer for your research and found it useful, please cite the manuscript:

Volker Herzig, David L. A. Wood, Felicity Newell, Pierre-Alain Chaumeil, Quentin Kaas, Greta J. Binford, Graham M. Nicholson, Dominique Gorse, Glenn F. King (2011) [ArachnoServer 2.0, an updated online resource for spider toxin sequences and structures](#), *Nucleic Acids Research* **39**, D653-D657 | [PubMed](#) | [Journal](#) |