Research Paper

BTDB - Database on Structural and Functional Information of Biotoxins

PERUMAL PERUMAL, RAJU GOWRIISHANKAR and SUNDARA BAALAJI NARAYANAN*
Structural Biology Lab, Department of Bioinformatics, School of Life Sciences, Bharathiar University, Coimbatore 641 046, India

Biotoxins, one of the hot diverse topics, enables us to identify the membrane proteins, ion channels and to predict the surface of cancer cells, consequently leading to discovery of novel drugs. The present work aims at providing the scientific community a quick review of the 477 toxic proteins from various species of snake, spider and scorpion, their structural features, putative functions and links to various online databases. In the present study each toxin protein's 3D structure was predicted by homology modeling and validated and the possible functions were annotated through computational methods. For the easy access of the theoretically modeled proteins and related data, database named BioToxin DataBase (BTDB) was created which is user friendly and the information collected from this database will have implications on further *in silico* and *in vitro* analysis on therapeutics, studies on membrane proteins and identification of ion channels.

This database can be accessed at the URL: http://b-u.ac.in/btdb/

Key Words: Biotoxin, Homology Modeling, RMSD, Minimized energy, Database, MySQL

Introduction

Biotoxins, the toxic peptides produced by animals, plants, fungi and bacteria, may be lethal or incapacitating (Patocka *et al.*, 2007) many of them representing a risk to human health and viability of organisms (Slater *et al.*, 2003). The biological mechanism of biotoxin as well as development of their toxic manifestation, clinical course and therapy pretensions varies for each toxin (Patocka *et al.*, 2007). Biotoxins are classified into three groups, based on their origin: plant biotoxins, microbial biotoxins and animal biotoxins. Animal biotoxins are highly potent short peptides of pharmacological importance. They are present in the venom of animals and used in ion channel's structure and functional studies, receptor studies, drug discovery and formulation of insecticides (Tan *et al.*, 2003).

Snakes, spiders and scorpions have evolved venom glands to produce potent toxins (Valentin and Lambeau, 2000). These are worth consideration because of the extreme diverse library of natural products they provide, that serves as a source of pharmacological tool (Fletcher *et al.*, 1997). The heterogeneous complex nature of the toxin with varying sizes (Escoubas *et al.*, 2000) and high potency act specifically on biological targets (Wang *et al.*, 2001).

Snake venoms produce swelling and hemorrhage to degenerative events and severe myolytic processes (Homma and Tu, 1971). They have been used in the drug development against malaria as it greatly reduces the efficiency of oocyst formation of malarial parasites (Zieler *et al.*, 2001). Saxatilin, a snake venom disintegrin, is known for inhibiting platelet aggregation, angiogenesis and melanoma pulmonary metastasis. It has been postulated that the anti-angiogenic and anti-metastatic activity of saxatilin might be due to the modulation

*Author for Correspondence: sundarabaalaji@gmail.com; sundarabaalaji@buc.edu.in; Tel: 91-422-2428285

of genes associated with migration and invasion of tumor cells (Kim *et al*., 2007).

Spider toxins contain a diverse array of peptides composed of high number of cysteine residues (Grishin, 1998). These toxins have been instrumental in determining the role and diversity of neuronal ion channels (Patocka *et al*., 2007) considering the following: the spider venom paralyzes its prey by modulating the activity of neuronal ion channels and receptors (Saez *et al*., 2010) and due to the similarity of ion channels between the natural invertebrate prey and vertebrates (Rash and Hodgson, 2002). Thus these peptide toxins represent a unique family of insecticidal neurotoxins that would be a key target for future site-direct mutagenesis and development as novel bioinsecticides (Nicholson and Graudins, 2002). Also, studies by Yan and Adams (1998) and Benli and Yigit (2008) showed that the venom of *L. carolinensis* and *A. labyrinthica* respectively, possess antimicrobial activity.

The venom of scorpions is a complex cocktail of homologous proteins with similar physicochemical properties but different pharmacological activities. Scorpion toxins target Na^+, K^+ and Ca^{2+} ion channels in membranes of excitable and non-excitable cells (Sands *et al*., 1989). These ion channel blockers could be potential immune suppressants for the treatment of autoimmune disorders such as rheumatoid arthritis, inflammatory bowel disease and multiple sclerosis (Jensen *et al*., 2002). Chlorotoxin found in the venom of *L. quinquestriatus* blocks the small conductance chloride channels (Debin and Strichartz, 1991) binding preferentially to glioma cells. Kievit *et al.* (2010) have reported that chlorotoxin labeled nanoparticles maybe used in the gene therapy against glioma.

Databases containing structural information have become indispensable for *in silico* structural analysis. Data for biotoxins are scattered across public databases providing sequences with limited descriptions and annotations on structure and function. As the gap between the amount of structural information and functional characterization widens, increasing efforts are being directed to the development of databases incorporating them. In view of the immense pharmacological and therapeutic importance of the biotoxins from snake, spider and scorpion group, we have undertaken the current study to determine their models along with their putative function and construct a database employing computational methods.

Methodology

Dataset Collection and Template Selection

A total number of 477 biotoxin sequences were retrieved for snake, spider and scorpion from the Uniprot database (http://www.uniprot.org/). Templates were selected through BLAST algorithm, with >40% sequence identity cut off (Altschul *et al*., 1990).

Tertiary Structure Prediction

Tertiary structure prediction by homology modeling was performed using Modeller 9V6 (Sali and Blundell, 1993).

Structure Validation

Structure validation for the modeled 3D structures were carried out using Structure Analysis and Verification Server (SAVES) (http://nihserver.mbi.ucla.edu/SAVES/) and PROCHECK analysis (Morris et al., 1992). GROMOS96 algorithm in SPDBV 4.0.1 was used to calculate the energy and for energy optimization (Guex and Peitsch, 1997).

Structural Features

Number of hydrogen bonds, number of helices, number of strands, number of turns and number of disulphide bridges were counted as structural features from the modeled 3D structures.

Function Prediction

The query protein sequences were analyzed by Protein Function prediction server (PFP) (http://kiharalab.org/web/pfp.php) to assign biochemical, molecular and cellular functions (Hawkins *et al*., 2006).

Database Construction

BTDB was created by using MySQL-5.1 as RDBMS, PHP for server side scripting and HTML for webpage designing. The tables were created with different datatypes to hold the structural and functional data of biotoxins, in the MySQL. PHP was used as an interface between HTML front end and MySQL back end to fetch data from database for user friendly display.

Results and Discussion

The BTDB is a user friendly database and provides essential structural information about the 477 biotoxic protein models from various species of snake, scorpion and spider. The 477 biotoxin sequences were retrieved randomly from Uniprot whose structures were not found in PDB. Only the template structures used in homology modeling were found in PDB. BTDB displays relevant information about biotoxic protein sequences, database accession ID, SwissProt ID, protein name, sequence length and sequence in fasta format. Databases on animal toxins (ATDB, http://protchem.hunnu.edu.cn/toxin/index.jsp) and food, pollutant toxins (T3DB, http://www.t3db.org/) exist which have details on name, molecular weight, classification, gene and protein sequence. In our database BTDB, structural and functional annotations on biotoxins are included, highlighting its uniqueness. The structural features contain the number of hydrogen bonds, number of helices, number of strands, number of turns and number of disulphide bonds. Based on the number of strands, helices, disulfide bridges in the predicted models, scope for the new classification of biotoxins is possible.

Annotation of Biotoxin in BTDB

Fig. 1 depicts the snapshot of the home page of BTDB. Structural and functional information of Accession ID BTSN070 (SwissProt ID: P25679) is given as follows. This snake toxin belongs to the

Fig 1: Snapshot showing the result page for the ID BTSN070. This page contains the details of predicted model, structure stability, structural information and the annotated function

species *Naja kaouthia* and it shares 64% of sequence similarity with the structure of Bungatoxin from *Bungarus Candidus* venom (PDB ID: 2JQP). Hence the 2JQP was used as a template to predict the structure. 29 hydrogen bonds, 5 strands, 6 turns and 5 disulphide bridges were present in the predicted model. No helices were found. The RMSD value of the predicted model was 0.419Å and the calculated energy was -3056.89 kcal/mol. The PROCHECK result describes that 81.5%, 16.7%, and 1.9% of residues were present in the most favorable region, additional allowed region and generously allowed region respectively, for the predicted BTSN070 model. The function for BTDB ID BTSN070 was annotated and the molecular functions have 16 predictions; the biological processes have 27 predictions and the cellular components have 17 predictions such as the ability to inhibit and to regulate the suitable receptors in molecular terms. From a biological perspective, the toxin takes part in transmission, cell death and in cellular component terms, the toxin have effect on postsynaptic membrane. These predictions were based on the probability value of sequence similarity.

### *Correlation Between RMSD, Sequence Similarity in BTDB*

The homology modeling concept is based on the conservative nature of motifs/domains. High percentage of identity can be correlated to a conservative nature and hence the sequence can be modeled more accurately. Stability of the modeled structure is directly proportional to the percentage of sequence similarity between the template and the query sequence. The modeled protein structures with RMSD value $\geq$ 4Å in the dataset have a positive energy indicating a low structural similarity with template. This implies quality of modeled proteins is poor. Hence refinement of the model is necessary in terms of energy to stabilize the structure. The biotoxins were also analyzed by the number of disulphide bridges present in the predicted models. The more disulphide bridges present, the stability of the biotoxin structure (Zhu *et al*., 2002; Possani *et al*., 1999) improves and exerts increased inhibitor activity as in the case of voltage gated ion channels

(Boubaker *et al*., 2004; Kharrat *et al*., 1997)

In the snake dataset, nearly 94% of structure has RMSD value below 4Å (Fig. 2a). These modeled structures have similarity greater than 70% with template and 80% of the modeled proteins have a low energy value and hence more structure stability (Fig. 3a). While in the spider dataset, the modeled structure showed significant deviation in sequence similarity with template. 19 sequences had less than 40% sequence similarity; 21 sequences had similarity in between 40%-60% and the remaining sequences had more than 60% similarity with template (Fig. 2b). This is also reflected in RMSD values and hence in the structural stability of spider biotoxins. Due to the radical deviation in the similarity, the modeled structures had a positive energy and refinement was carried out yielding a different conformation of the same model with low acceptable energy level (Fig. 3b). In the scorpion dataset, the modeled structures can be considered more reliable upon validation as nearly 96% of sequences with template similarity $\geq$ 60% were found to have RMSD value < 4Å (Fig. 2c). Fig. 3c shows the drastic deviation in sequence similarity and wide range of minimized energy values. 95% of the predicted models have template similarity $\geq$50% was stable in scorpion dataset These values establish a direct relation between RMSD and sequence similarity which substantiates the stability of the theoretically modeled biotoxins. Thus, the BTDB contains validated and stable theoretical models. This makes it easier for the structure determination of biotoxins through *in vitro* techniques like X-ray crystallography.

Conclusion

The structural data available in our database is more reliable and the models have structural stability. The BTDB would soon find utility in the scientific community for a quick review of the biotoxins and their structures. From the BTDB database, researchers can use the structures of biotoxins in the field of cancer studies, identifying ion channels, blocking the membrane proteins through *in silico* structural and functional analysis. The Biotoxin database will be updated in future with potent biotoxins from animals,
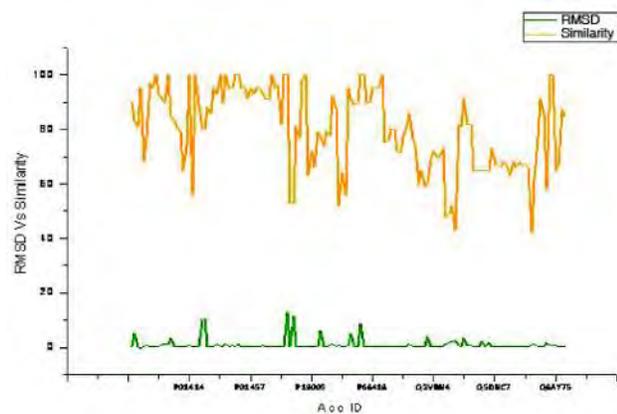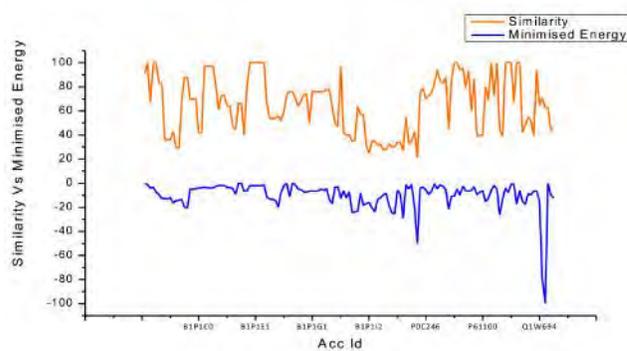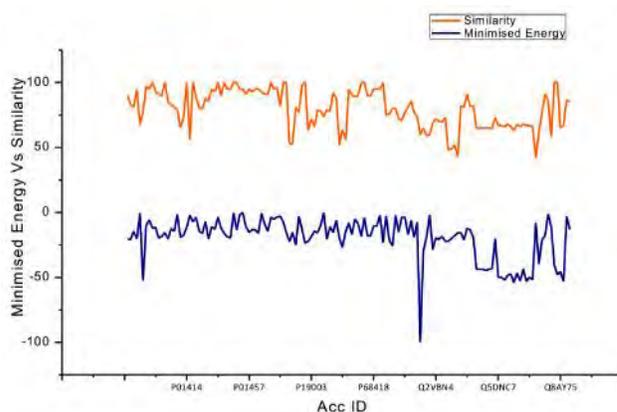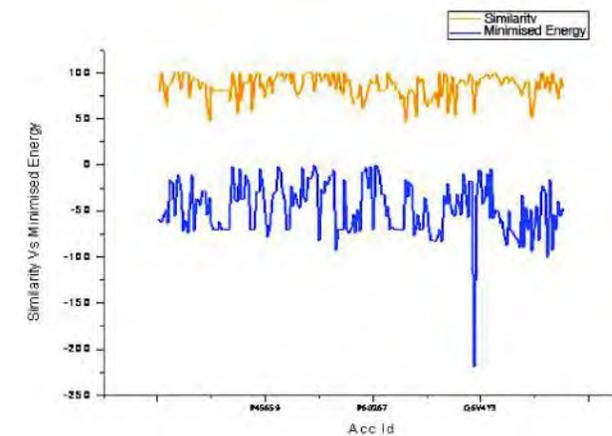
(a)
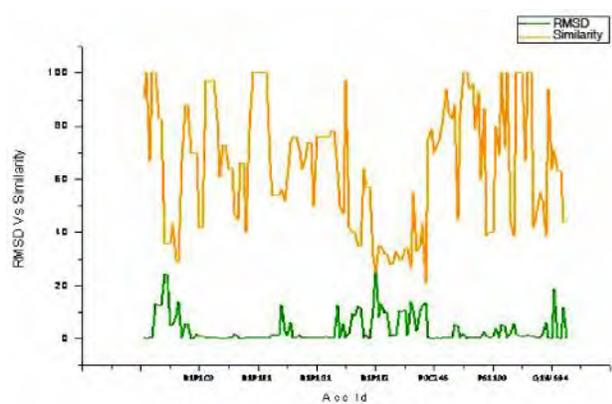
(b)
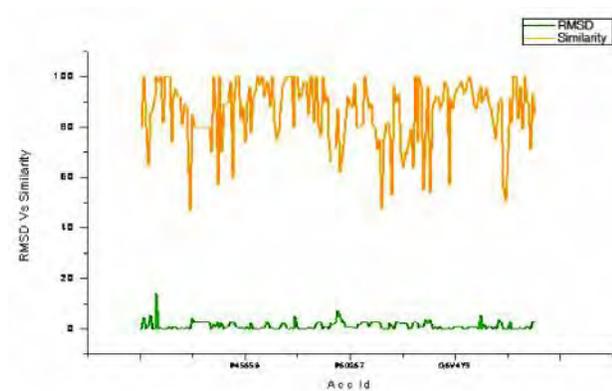
(c)

Fig. 2: **Graph showing the RMSD Vs Similarity values for (a) Snake Biotoxin, shows 94% of model's RMSD values fall below 4Å. (b) Spider Biotoxin, shows drastic deviation of RMSD values due to lack of sequence similarities. (c) Scorpion Biotoxin, shows 96% of sequences with template identity more than 60% were found to have RMSD value of less than 4Å**

(a)

(b)

(c)

Fig. 3: **Graph showing the Similarity Vs Minimized energy values for (a) Snake Biotoxin, shows 80% of the modeled proteins have low energy value. (b) Spider Biotoxin, describes the energy minimization values. (c) Scorpion Biotoxin. Nearly 95% sequences had low energy due to high sequence similarity with templates**

plants and microbes. Vital data on structural information like molecular property will be added that will serve as a platform for drug discovery.

Acknowledgement

References

Altschul S F, Gish W, Miller W, Myers E W and Lipman D J (1990) Basic local alignment search tool *J Mol Biol* **215** 403-410

Benli M and Yigit N (2008) Antibacterial Activity of Venom from Funnel Web Spider *Agelena labyrinthica J Venom Anim Toxins incl trop Di* **14** 641-650

DeBin J A and Strichartz G R (1991) Chloride channel inhibition by the venom of the scorpion *Leiurus quinquestriatus Toxicon* **29** 1403-1408

Escoubas P, Diochot S and Corzo G (2000) Structure and pharmacology of spider venom neurotoxins *Biochemie* **82** 893-907

Fletcher J I, Chapman B E, Mackay J P, Howden M E and King G F (1997) The structure of versutoxin (deltaatracotoxin-Hv1) provides insights into the binding of site 3 neurotoxins to the voltage-gated sodium channel *Structure* **5** 1525-1535

Grishin E V (1998) Black widow toxins, The present and the future *Toxican* **36** 693-1701

Guex N and Peitsch M C (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling *Electrophoresis* **18** 2714-2723

Hawkins T, Luban S and Kihara D (2006) Enhanced Automated Function Prediction Using Distantly Related Sequences and Contextual Association by PFP *Protein Sci* **15** 1550-1556

Homma M and Tu A T (1971) Morphology of local tissue damage in experimental Snake envenomation *Br J Exp Pathol* **52** 538-542

Jensen B S, Hertz M, Christophersen P and Madsen L S (2002) The Ca2+-activated K+ channel of intermediate conductance:a possible target for immune suppression *Expert Opin Ther Targets* **6** 623-36

Kharrat R, Mansuelle P, Sampieri F, Crest M, Oughideni R, Rietschoten J V, Martin-Eauclaire M F, Rochat H and Ayeb M E (1997) Maurotoxin, a four disulfide bridge toxin from Scorpio maurus venom: purification, structure and action on potassium channels *FEBSLett* **406** 284-290

Kievit F M, Veiseh O, Fang C, Bhattarai N, Lee D, Ellenbogen R G and Zhang M (2010) Chlorotoxin labeled magnetic nanovectors for targeted gene delivery to glioma *ACSnano* **4** 4587-94

Kim D S, Jang Y, Jeon O and Kim D (2007) Saxatilin, a Snake Venom Disintegrin, Suppresses TNF-α-induced Ovarian Cancer Cell Invasion *Journal of Biochemistry and Molecular Biology* **40** 290-294

Mahjoubi-Boubaker B, Crest M, Khalifa R B, Ayeb M E and Kharrat R (2004) Kbot1, a three disulfide bridges toxin from Buthus occitanus tunetanus venom highly active on both SK and Kv channels *Peptides* **25** 637-645

Morris A L, MacArthur M W, Hutchinson E G and Thornton M J (1992) Stereochemical quality of protein structure coordinates *Proteins* **12** 345-364

Nicholson G M and Graudins A (2002) Spiders of medical importance in the Asia-Pacific: atracotoxin, latrotoxin and related spider neurotoxins *Clin Exp Pharmacol Physiol* **29** 785-794

Patocka J, Hon Z, Streda L, Kuca K and Jun D (2007) Biohazards of Protein Biotoxins *Defence Science Journal* **57** 825-837

Possani L D, Selisko B and Gurrola G B (1999) Structure and function of scorpion toxins affecting K+-channels *Perspectives in Drug Discovery and Design* **15** 15-40

Rash L D and Hodgson W C (2002) Pharmacology and biochemistry of spider venoms *Toxicon* **40** 225-254

Saez N J, Senff S, Jensen J E, Er S Y, Herzig V, Rash L D and King G F (2010) Spider-Venom Peptides as Therapeutics *Toxins* **2** 2851-2871

Sali A and Blundell T L (1993) Comparative protein modeling by satisfaction of spatial restraints *J Mol Biology* **234** 779-815

Sands S B, Lewis R S and Cahalan M D (1989) Charybdotoxin Blocks Voltage-gated K + Channels in Human and Murine T Lymphocytes *J Gen Physiol* **93** 1061-1074

Slater L N and Greenfield R A (2003) Biological toxins as potential agents of bioterrorism *J Okla State Med Assoc* **96** 73-76

Tan P T J and Khan A M V (2003) Brusic Bioinformatics for venom and toxin sciences) *Briefings in bioinformatics* **4** 5362

Valentin E and Lambeau G (2000) What can venom phospholipase A2 tell us about the functional diversity of mammalian secreted phospholipase A2 *Biochimie* **82** 815-831

Wang X, Connor M, Smith R, Maciejewski M W, Howden M E, Nicholson G M, Christie M J and King G F (1999) Structure-function studies of omega-atracotoxin, a potent antagonist of insect voltage-gated calcium channels *Eur J Biochem* **264** 488-494

Yan L and Adams M E (1998) Lycotoxins, antimicrobial peptides from venom of the wolf spider *Lycosa carolinensis J Biol Chem* **273** 2059-2066

Zhu Q, Liang S, Martin L, Gasparini S, Ménez A and Vita C (2002) Role of disulfide bonds in folding and activity of leiurotoxin I: just two disulfides suffice *Biochemistry* **41** 11488-11494

Zieler H, Keister D B, Dvorak J A and Ribeiro J M C (2001) A snake venom phospholipase A2 blocks malaria parasite development in the mosquito midgut by inhibiting ookinete association with the midgut surface *The Journal of Experimental Biology* **204** 157-4167.