In Silico Structural and Functional Analysis of Phospholipase D from Brown Recluse Spider (*Loxoceles reclusa)* Venom

A DISSERTATION SUBMITTED TO BRAC UNIVERSITY IN PARTIAL FULFILMENT OF THE REQUIRMENTS FOR THE BACHELORS OF SCIENCE IN BIOTECHNOLOGY

Submitted by-

AQIEF AFZAL

ID: 12236007

June 2017

Biotechnology Program

Department of Mathematics and Natural Sciences

BRAC University

Bangladesh

Dedicated

To my

Beloved Parents

DECLARATION

I hereby declare that the research work embodying the results reported in this thesis entitled "*In Silico* Structural and Functional Analysis of Phospholipase D from Brown Recluse Spider (*Loxoceles reclusa*) Venom" submitted by the undersigned has been carried out under the supervision of Dr. M. Mahboob Hossain, Professor, Department of Mathematics and Natural Sciences, BRAC University, Dhaka. It is further declared that the research work presented here is original and has not been submitted to any other institution for any degree or diploma.

Candidate:

_____

AQIEF AFZAL

Certified:

_____

Dr. M. Mahboob Hossain

Supervisor

Professor

Department of Mathematics and Natural Sciences

BRAC University, Dhaka

Acknowledgment

It's my honor to write this acknowledgment to express my gratefulness to people who have helped me throughout this research work.

First of all, I would like to thank The Almighty Allah for successfully accomplishing my thesis work.

I am grateful to Professor A. A. Ziauddin Ahmad, Chairperson, Department of Mathematics and Natural Sciences, BRAC University, for allowing me to pursue my undergraduate studies in the Department of MNS and for his constant guidance throughout my entire period of study in the department.

My humble respect goes to honorable Professor Naiyyum Chowdhury, Coordinator, Biotechnology and Microbiology, Department of Mathematics and Natural Sciences, BRAC University, for his inspiration and prudent advice and also for giving me the opportunity to gain significant experience.

I would like to convey my heightened appreciation thanks to Dr. M. Mahboob Hossain, Professor, Department of Mathematics and Natural Sciences, BRAC University, for his valuable suggestions, unforgettable help and soft behavior also affectionate guidance and for giving me the opportunity to work under her supervision.

I also express my special thanks to all my respected teachers at the Department of Mathematics and Natural Sciences, BRAC University, for their academic counsel and encouragement.

I would particularly like to express my deepest thanks to Ms. Abira Khan, Faculty, Department of Genetic Engineering and Biotechnology, University of Dhaka, for her hearty, incessant cooperation and encouragement throughout the study. I gratefully acknowledge her for her advice and crucial contribution throughout my research and thesis writing periods.

In addition to the above, I would like to thank Ms. Eusra Mohammad, Lecturer, Department of Mathematics and Natural Sciences, BRAC University, for her help in giving a proper shape to the thesis.

Also I thank deeply from my heart to my lovely and helpful seniors Rhedia Tehrin Proma and Mayaboti Akbar Cinderella, who provided constant encouragement, sound advice, good company, and lots of good ideas throughout the study and also I would like to thank my friends and juniors for their enthusiastic inspiration and company during my thesis work.

The whole credit of my achievements during the research work goes to my parents. It was their cordial effort that acquainted me to the way of research.

AQIEF AFZAL
ID: 12236007
Department of Mathematics and Natural Sciences
BRAC University, June 2017.

ABSTRACT

Spiders are an interesting class of insects - omnipresent in almost in all environment and vet, mostly remains unnoticed. They are carnivores in nature, relying on their own venom to neutralize their prey. There are many poisonous spider species but little is heard about spider bites in Bangladesh. The Brown Recluse, *Loxosceles reclusa* is one of the many spider species found in Bangladesh. It has deadly necrotic venom several components of which are medically significant. This project aims to study the venom composition of Brown Recluse and analyze the structure and function of a specific venom component. This *In Silico* project will help in narrowing wet lab research to particular venom proteins with medically useful attributes and thus reduce waste of wet lab resources and make future projects cost effective. Various online bioinformatics resources were utilized to predict the composition, structure and functional characteristics of the spider venom components. In the initial phase, a thorough literature review was carried out to elucidate the venom composition of the Brown Recluse spider. Then the desired target protein, Phospholipase D, was selected based on its potential therapeutic application. The protein was subjected to sequence based homology search to find out with other proteins sequences with best matches to it. These protein sequences were analyzed using global alignment to look for stretches of conserved regions within them. Predicted signal peptide is 20 a.a.long with cleavage between Asn^{18} and His^{19}. In the later part, a phylogenetic tree was constructed based on their sequences to see their evolutionary divergence in respect to each other. It predicted that the target protein sequence was evolutionarily quite divergent and was placed in a separate sub group from the other proteins despite being similar to them in sequence. The target protein sequence was also analyzed using protein motif finding software to look for functional motifs. The protein motif finding software's were able to find some motif or binding sites. Namely, three binding sites were seen, GDPD, PLC and Signal peptides. Lastly, the target protein's 3D structure was constructed based on template structures of other proteins with similar sequences. The result predicted that the protein is almost globular in shape with a core formed by beta sheets on the inside and alpha helices forming the outer covering. This study shows that the structure of the target proteins quite unique- having a highly conserved sequence and yet being evolutionarily distant from other similar proteins from related spider species.

CONTENT

ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
GO	Gene Ontology
NCBI	National Center for Biotechnology Information
kDa	Kilo Dalton
PLD	Phospholipase D
LPC	lysophosphatidylcholine
C1	Pceramide-1-phosphate
LPA	lysophosphatidic acid

Introduction

1. Introduction:

1.1 General Overview:

"The parable of those who take protectors other than Allah is that of the spider, who builds (to itself) a house; but truly the flimsiest of houses is the spider's house if they but knew Surah Ankaboot: 41; (Al-Quran)."

Spiders are carnivorous arthropods and consume a large number of preys. They have unique habitat in almost all the environments. Spiders serve as buffers that limit the initial exponential growth of prey populations. The predatory spiders are classified into five major groups based on their foraging style. Prey searching ability, wide host range, ease in multiplication and polyphagous nature make them potential predators in biological pest suppression. Spiders can survive in most environments and they have great biodiversity.

Spiders belong to the order **Araneae**, class ***Arachinidae*** and are members of the phylum **Arthropoda**, the largest assemblage of the animal with jointed legs and a hard exoskeleton. They are the largest group of *arachinids* comprising more than 30,000 species distributed over 60 families worldwide [1].

Fifty spider species in 24 genera and 13 families have been reported till date in Bangladesh of which 18 are endemic to the country. In spite of a large geographical area, there have been very few studies on spiders in this country. About 50% of the spiders reported in the South Asian region have been described from India [2]. Recently there has been a spurt in spider research by Bangladeshi entomologists.

Spiders play a very important role in balancing the biodiversity. Spiders are a good pest control in crop fields and many drugs are made from the venom of spiders [3].

Table 1.1

Spider species of Bangladesh

Family Name	**Genus Name**
ARANEIDAE	*Araneus*
	Arglope
	Cyclosa
	Cyrtopthora
	Gea
	Neoscona
CLUBIONIDAE	*Clubiona*
CORINNIDAE	*Castianeira*
	Trachelas
LINYPHIIDAE	*Callitrichia*
LIOCRANIDAE	*Sphingius*
LYCOSIDAE	*Lycosa*
	Pardosa
MITURGIDAE	*Cheiracanthium*
PHILODROMIDAE	*Tibellus*
PISAURIDAE	*Hygropoda*
SALTICIDAE	*Marpissa*
	Phidippus
SALTICIDAE	*Plexippus*
	Zeuxippus
TETRAGNATHIDAE	*Leucauge*
	Dyschiriognatha
	Tetragnatha
THERAPHOSIDAE	*Chilobrachys*
THOMISIDAE	*Runcinia*

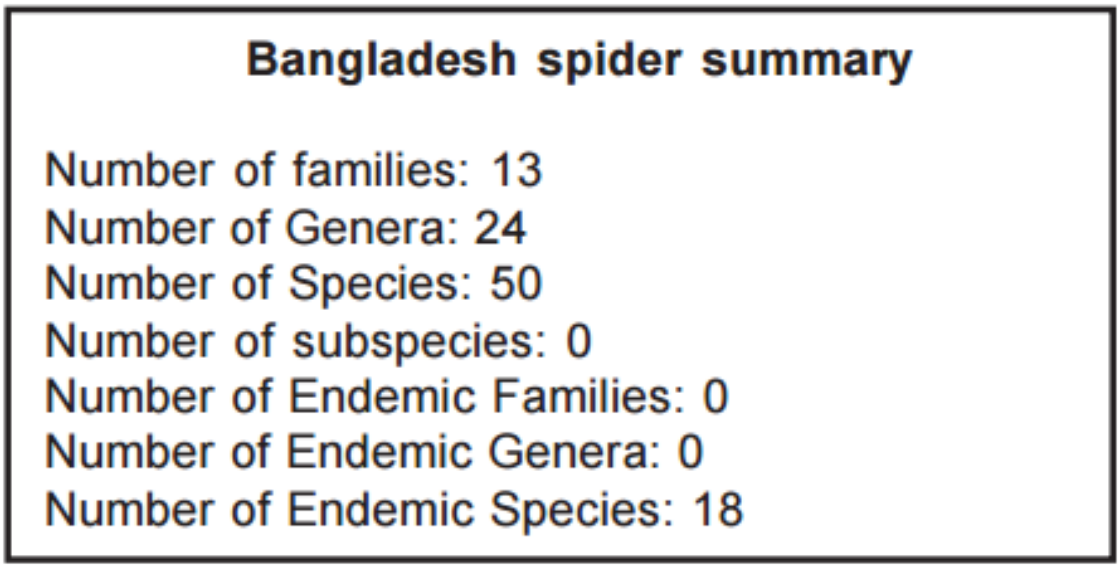

Table 1.2 Summary of Bangladeshi Spider species

The Brown Recluse, *Loxosceles recluse* is one of the many spider species in Bangladesh. The reason behind choosing this species is that its venom has shown to have several components with potential therapeutic attributes.

1.2 Aim and objectives of the Thesis

The project had the following aims and objectives:

- Study the venom composition of the **Brown Recluse** spider**.**
- Analyze the structure of the Phospholipase D present in the Brown Recluse spider venom.
- Find the homologous regions (with Phospholipase D from other spider species) and conserved regions of the sequence.
- Find functional motifs and annotate the sequence.
- Infer phylogenetic relationship with homologous sequence.
- Predict the 3D structure and substrate binding site of the sequence.

1.3 BROWN RECLUSE (Loxosceles reclusa)

The brown recluse, Loxosceles reclusa, **Sicariidae** (formerly placed in the family "**Loxoscelidae**") is a recluse spider with necrotic venom and is one of two spiders (the other being the black widow) with medically significant venom in North America. Their bite, like various other recluse bites, sometimes requires medical attention [4].

Brown recluse spiders are usually between 6 and 20 millimeters (0.24 and 0.79 in) but may grow larger. While typically light to medium brown, they range in color from whitish to dark brown or blackish gray. The cephalothorax (the front segment of each spider) and abdomen are not necessarily the same color. These spiders usually have markings on the dorsal side of their cephalothorax, with a black line coming from it that looks like a violin with the neck of the violin pointing to the rear of the spider, resulting in the nicknames fiddle back spider, brown fiddler, or violin spider [5].

The Brown Recluse Spider is a reclusive creature that seeks and prefers seclusion. The Brown Recluse Spider and ten additional species of Loxosceles are native to the United States. In addition, a few non-native species have become established in limited areas of the country. The Brown Recluse Spider is found mainly in the central Midwestern states southward to the Gulf of Mexico. A related species, the brown violin spider is found in Hawaii [6].

FIG. 1.3: THE BROWN RECLUSE

1.4 Taxonomy of the Brown Recluse [7]

Kingdom Animalia (Animals)

Phylum Arthropoda (Arthropods)

Subphylum Chelicerata (Chelicerates)

Class Arachnida (Arachnids)

Order Araneae (Spiders)

Infraorder Araneomorphae (True Spiders)

Family Sicariidae (Recluse Spiders)

Genus Loxosceles (Recluse Spiders)

Species Loxosceles reclusa (Brown Recluse)

1.5 Habitat of the Brown Recluse:

Brown recluse spiders build asymmetrical (irregular) webs that frequently include a shelter consisting of a disorderly thread. They frequently build their webs in woodpiles and sheds, closets, garages, plenum spaces, cellars, and other places that are dry and generally undisturbed. When dwelling in human residences they seem to favor cardboard, possibly because it mimics the rotting tree bark which they inhabit naturally. They have also been encountered in shoes, inside dressers, in bed sheets of infrequently used beds, in clothes stacked or piled or left lying on the floor, inside work gloves, behind baseboards and pictures, in toilets, and near sources of warmth when ambient temperatures are lower than usual. Human-recluse contact often occurs when such isolated spaces are disturbed and the spider feels threatened. Unlike most web weavers, they leave these lairs at night to hunt. Males move around more when hunting than the females, which tend to remain nearer to their webs. The spider will hunt for firebrats, crickets, cockroaches, and other soft-bodied insects [8].

1.6 Spider Venom Composition

More than 46,000 species of spider creepy crawl across the globe, on every continent except Antarctica. Each species produces venom composed of an average of 500 distinct toxins, putting the conservative estimate of unique venom compounds at more than 22 million. This staggering diversity of venoms collectively referred to as the venom, has only begun to be explored.

Among the handful of brave scientists studying spider venom are Greta Binford at Lewis and Clark College in Portland, Oregon, and Jessica Garb at the University of Massachusetts at Lowell. Both of these researchers analyze the protein structures of various venom chemicals in search of clues that can explain why some are lethal (see side bar below), while the vast majority are thought to be relatively harmless. The scientists also use molecular biology tools to compare the genomes of spiders that have extremely noxious venoms, including the black widow and the brown recluse, to those of spiders with non-poisonous venoms, such as the house spider.

The data being collected by Garb and Binford and their colleagues have the potential to increase our understanding of the evolution of spider venom and hopefully contribute to the discovery of new medicines, anti-venoms and insecticides. In addition, because many of the deadlier spider venoms produce their toxic effects by over stimulating the production of brain signaling molecules, this research may uncover novel tools for neuroscience research [9].

The majority of spiders possess neurotoxin venom. These neurotoxins are multicomponent but contain three main groups of toxic compounds:

1. Low M_r polyamines (M_r less than 1000)
2. Polypeptides (M_r 3000-10,000)
3. High M_r proteins (M_r more than 10,000)

Other venom components include inorganic ions and salts, free acids (eg. lactic acid), glucose, nucleic acids, free amino acids and biogenic amines. The exact role of these is unknown but they are thought to aid in the stability, delivery and effectiveness of the toxins [10].

Spider venom peptide toxins are active as myotoxins and neurotoxins and are potassium, sodium, and calcium ion channel active compounds (3, 4, 5, and 6). Lower molecular weight acylpolyamines are paralytic and composed of an aromatic portion, amino acids, and polyamines; indole and hydroxybenzene are linked to spermine and other polyamines by an acyl group or amino acid. These compounds are major components of many spider venoms and have been called "glutamate receptor antagonists". Hyaluronidase (also known as "spreading factor"), nucleotide-like adenosine triphosphate (ATP) and citrate are some other major spider venom components. Hyaluronidase and citrate (anticoagulant, chelator, and buffer) can be proposed as active venom components, and recently extracellular ATP has been reported as toxic to certain cells [11]

The dermo necrotic toxin of *Loxosceles*, 32 kDa, is reported to be a Sphingomyelinase enzyme, and recent reports have indicated the development of effective anti-venom is possible. The action of *Loxosceles* toxin is not well understood [12].

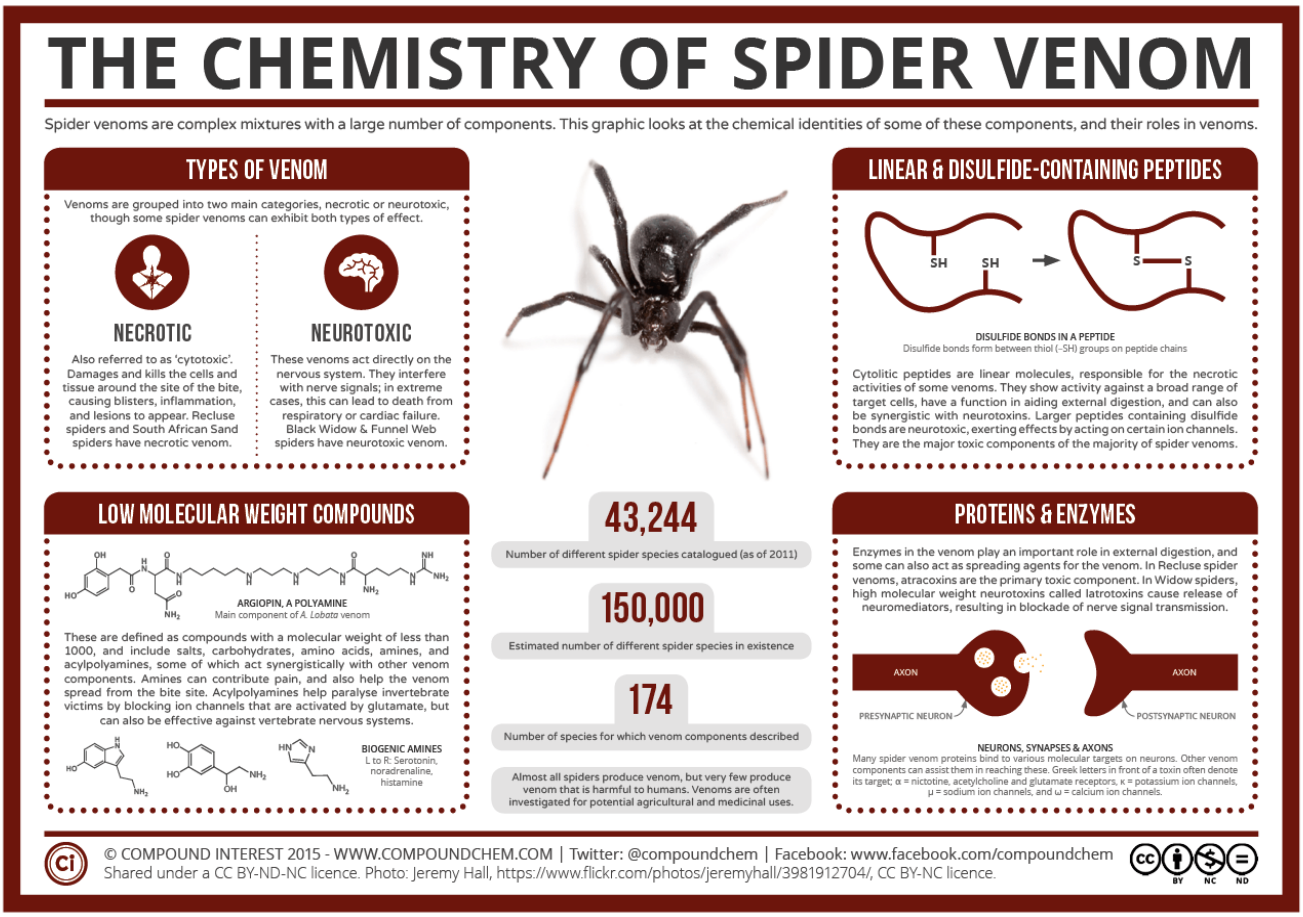

FIG 1.6. The whole chemistry behind the spider venom

An example of a spider venom neurotoxic protein is α-latrotoxin from the black widow spider. It is highly toxic to vertebrates and causes massive neurotransmitter release. The active enzyme in brown recluse spider venom is Sphingomyelinase D or Phospholipase D which causes the degradation of cell membranes and the development of painful lesions.

Table 1.6.1 Various components of the spider venom [13]

Venom Component	Mode of Action	Families Using This Toxin
Acylpolyamines	Non-competitive inhibition at glutamate receptors, leading to paralysis	Araneidae, Nephilidae, and some other families
Peptides with disulfide (-s-s-) bridges twisting molecule into a knot	Uncontrolled stimulation or inhibition at sodium, potassium or calcium ion channels in nerve endings	Many families but especially the Hexathelidae, Actinopodidae, Theraphosidae, Agelenidae
Neurotoxic proteins	Exhaustive release of neuromodulators at nerve endings, causing intense pain	Theridiidae, especially some Latrodectus and Steatoda species
Linear cytolytic peptides	Damage to cell walls in a victim's tissues, causing the cells to break open and die	Zodariidae, Lycosidae, Oxyopidae, Araneidae, Agelenidae
Histolytic enzymes	Tissue ulceration by 'ungluing' of tissues, notably involving sphingomyelinase D	Sicariidae, especially Loxosceles species
Digestive enzymes, especially hyaluronidase and collagenase	Dissociation of tissue cells to promote extracorporeal digestion	Many families including the Desidae, Lycosidae, Sicariidae, Theridiidae, and Sparassidae
Small acids and amines	Pain-producing and some other actions	Many families notably including the Salticidae and Hexathelidae

Table 1.6.2. Spider Venom Component as Drugs

Table 8.1 Patents detailing potential therapeutic applications of spider-venom peptides filed during the period 1989–2014.

Patent number	*Year*	*Assignee*	*Peptide name*	*Spider species*	*No. of residues*	*SS bonds*	*Target/mode of action*	*Disease/ application*
US 8 664 179	2014	CNRS & Universite de Nice Sophia Antipolis	Psp3Tx1	*Paraphysa* spp.	28	3	Inhibitor of T-type voltage-gated calcium channel $Ca_V3.2$	Pain, hypertension, and prostate cancer
WO2014016673 A1	2014	Purdue Pharma L.P.	C-terminally modified protoxin-II	*Thrixopelma pruriens*	30	3	Inhibitor of voltage-gated sodium channels, particularly $Na_V1.7$	Na_V channel disorders, especially pain
WO 2013173706 A2	2013	Janssen Biotech Inc.	Huwentoxin-IV analogs	*Haplopelma schmidti*	39	3	Inhibitor of voltage-gated sodium channel $Na_V1.7$	Pain
US 8 399 026	2013	Alomone Preclinical Ltd	Peptides A, B	*Pterinochilus* spp.	34, 29	3, 3	Inhibitor of voltage-gated sodium channels $Na_V1.3$ and $Na_V1.8$	Pain
			Peptides C, D	*Haplopelma lividum*	35, 33	3, 3		
			Peptide E, VSTX3	*Grammostola* spp.	34, 34	3, 3		
US 8 383 162 WO2008061329 A2	2013	Universidade Federal de Minas Gerais	Phα1β	*Phoneutria nigriventer*	55	6	Inhibitor of N-type voltage-gated calcium channel $Ca_V2.2$	Pain
WO2012125973 A2	2012	Amgen Inc.	GpTx-1 (ω-TRTX-Gr2a)	*Grammostola porteri*	34	3	Potent inhibitor of $Na_V1.3$ and $Na_V1.7$. Weak inhibitor of $Ca_V3.1$	Pain
US 7 892 764	2011	Legacy Emmanuel Hospital & Health Center	PcTx1 (π-TRTX-Pc1a)	*Psalmopoeus cambridgei*	40	3	Selective blocker of ASIC1a	Epilepsy (seizure suppression)
WO2009094742 A1	2009	Universidade Federal de Minas Gerais	Tx2-6 (δ-CNTX-Pn2a)	*Phoneutria nigriventer*	48	5	Inhibits inactivation of Na_V channels and increases NO release in the corpus cavernosum	Erectile dysfunction

1.7 Spider Venom Enzyme Components and their Mode of Actions:

Table 1.7 The main enzymes of spider venom

Type	Name	Origin
Oxydoreductases	dehydrogenase lactate	Elapidae
	L-amino-acid oxidase	All species
	Catalase	All species
Transferases	Alanine amino transferase	
Hydrolases	Phospholipase A_2	All species
	Lysophospholipase	Elapidae, Viperidae
	Acetylcholinesterase	Elapidae
	Alkaline phosphatase	Bothrops atrox
	Acid phosphatase	Deinagkistrodon acutus
	5'-Nucleotidase	All species
	Phosphodiesterase	All species
	Deoxyribonuclease	All species
	Ribonuclease 1	All species
	Adenosine triphosphatase	All species
	Amylase	All species
	Hyaluronidase	All species
	NAD-Nucleotidase	All species
	Kininogenase	Viperidae
	Factor-X activator	Viperidae, Crotalinae
	Heparinase	Crotalinae
	α-Fibrinogenase	Viperidae, Crotalinae
	β-Fibrinogenase	Viperidae, Crotalinae
	α-β-Fibrinogenase	Bitis gabonica
	Fibrinolytic enzyme	Crotalinae
	Prothrombin activator	Crotalinae
	Collagenase	Viperidae
	Elastase	Viperidae
Lyases	Glucosamine ammonium lyase	

1.8 Effects of spider bite [14]

When the spider bites people, they may feel a sharp sting or nothing at all. Pain usually develops within the first several hours after being bitten and may become severe. Children may have more serious reactions.

Symptoms may include:

- Chills
- Itching
- General ill-feeling or discomfort
- Fever
- Nausea
- Reddish or purplish color in a circle around bite
- Sweating
- Large sore (ulcer) in the area of the bite

Rarely, these symptoms may occur:

- Coma
- Blood in urine
- Yellowing of the skin and whites of the eyes (jaundice)
- Kidney failure
- Seizures

1.9 Analysis of Brown Recluse Phospholipase D protein

Phospholipase D [15]

Phospholipase D catalyses the hydrolysis of the phosphodiester bond of glycerophospholipids to generate phosphatidic acid and a free head group. Phospholipase D activities have been detected in the simple to complex organisms from viruses and bacteria to yeast, plants, and mammals. Although enzymes with broader selectivity are found in some of the lower organisms, the plant, yeast, and mammalian enzymes are selective for phosphatidylcholine. The two mammalian Phospholipase D isoforms are regulated by protein kinases and GTP binding proteins of the ADP-ribosylation and Rho families. Mammalian and yeast Phospholipase D are also potently stimulated by phosphatidylinositol 4,5-bisphosphate.

PLD-catalyzed hydrolysis has been proposed to occur in two stages via a "ping-pong" mechanism. In this scheme,

- The histidine residues of each HKD motif successively attack the phospholipids substrate.
- Functioning as nucleophiles, the constituent imidazole moieties of the histidines form transient covalent bonds with the phospholipids, producing a short-lived intermediate that can be easily hydrolyzed by water in a subsequent step.

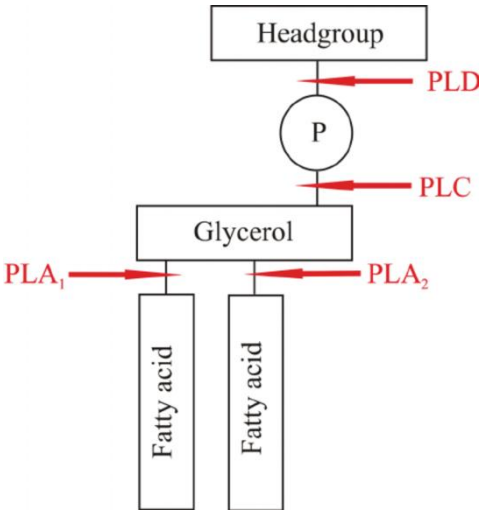

FIG 3: Mode of Action of Phospholipase D

Phospholipase D in the spider venom:

Envenomations by brown spiders in the genus *Loxosceles* can induce a disease state called loxoscelism in mammalian tissue. Cutaneous loxoscelism can result in ulcer formation, edema, and dermonecrosis at the site of envenomation. Some envenomations result in systemic loxoscelism involving hemolysis, circulatory shock, intravascular coagulation, renal failure and even death.

Phospholipase D (PLD) toxins in the venom are the primary agents responsible for loxoscelism. Toxins purified from venom or recombinant sources elicit the full pathology of loxoscelism when injected into animal models. Multiple PLD isoforms and homologs are expressed in venoms throughout the spider family Sicariidae, which includes the genera *Loxosceles* and *Sicarius*. The gene family comprising these toxins has been named SicTox to reflect this phylogenetic distribution.

PLD toxins from Loxosceles bind to mammalian cell surfaces and have enzymatic activity against common phospholipids in mammalian tissue, including lysophosphatidylcholine (LPC) and sphingomyelin (SM). The most common activity assay for these enzymes detects PLD activity by monitoring choline release from the substrate. Liberation of choline from SM or LPC is commonly assumed to result from substrate hydrolysis, giving either ceramide-1-phosphate (C1P) or lysophosphatidic acid (LPA), respectively, as a second product. C1P and LPA are thought to contribute to the pathology of loxoscelism by some as the yet undetermined mechanism but definitive evidence for the formation of these phosphate-containing products is lacking. 31P NMR spectroscopy and mass spectrometry to directly observe the formation of the phosphate-containing products from the action of a recombinant Loxosceles PLD toxin, as well as diverse whole venom samples, on the substrates SM and LPC. Instead of the hydrolytic products C1P and LPA, respectively, we observe exclusive formation of cyclic phosphate products resulting from intramolecular transphosphatidylation [16]

Phospholipase D in *Loxoceles*

The brown recluse spider, *Loxosceles reclusa*, and other species of the *Loxosceles* genus also express PLD-like SMases, called SMase Ds with significant sequence homology to the *Corynebacterium* and *Arcanobacterium* SMase D that catalyzes manganese-dependent acid-base hydrolysis of SM to ceramide-1-phosphate. These enzymes are the major component of spider venom and are responsible for the dermonecrosis, haemolysis, dysregulated neutrophil activation, and other toxic physiological responses to a spider bite. *In vitro* characterization of recombinant SMase Ds demonstrates that these enzymes can be divided into two classes based on biochemical activity. Class I enzymes, in addition to SMase activity, also efficiently hydrolyze lysophospholipids in a PLD-like fashion to generate LPA. As such, it has been proposed this enzyme be called a PLD, a less specific title that encompasses all activities characterized to date. LPA triggers signaling cascades in the host organism including platelet aggregation and inflammatory response. Lyso PLD activity is likely the cause for some of the lethal effects of *Loxosceles* venom. Class II enzymes exhibit SMase activity and exhibit decreased activity towards phospholipids. Subsequently, other spider and snake venoms have been characterized as having PLD or SMase D activity. The crystal structures of class I and class II SMase enzymes from *Loxosceles* reveal differences in the catalytic cleft that explain observed differences in substrate selectivity [17].

Materials and Methods

2. Materials and Methods [18]:

2.1 Protein Sequence

The target protein sequence (FASTA format) was searched for in GenPept database.

2.2 Homology

Homology (sequence/structure similarity) among protein sequence implies their functional relatedness. If 2 sequences are significantly homologous, it implies they perform the same function or belong to the same protein family. The following software was used to study homology of Brown Recluse Phospholipase D precursor proteins.

PSI-BLAST

PSI-BLAST (Position specific iterated basic local alignment search tool) is a provided on NCBI.

- Based on BLOSUM62 scoring matrix
- Finds similarities between distantly related proteins.
- Help to identify whether a new protein is a member of an already characterized protein family.
- By repeatedly (iteration) searching and adjusting scoring matrix for alignments, it brings out sequences significantly homologous to a target sequence.
- It finds out regions of local alignments (homologous stretch) among proteins.

COBALT

COBALT does progressive multiple alignments of protein sequences. The alignment is aided by a collection of pair wise constraints derived from conserved domain database, protein motif database, and local sequence similarity using RPS-BLAST, BLASTP, and PSI-BLAST, respectively. Computation time is reduced by forming clusters of sequences that share a large number of common words and finding conserved domains and motif matches for only one sequence per cluster.

2.3 Protein motif

Proteins are generally comprised of one or more functional regions, commonly termed domains. The presence of different domains in varying combinations in different proteins gives rise to the diverse repertoire of proteins found in nature. Identifying the domains present in a protein can

- Provide insights into the function of that protein.
- If the sequence of an unknown protein is too distantly related to any protein of known structure to detect its resemblance by overall sequence alignment, it can be identified by the occurrence in its sequence of a particular cluster of residue types which is variously known as a pattern, motif, signature, or fingerprint. These motifs arise because of particular requirements on the structure of specific region(s) of a protein which may be important, for example, for their binding properties or for their enzymatic activity.

The Brown Recluse Spider Venom sequence was analyzed for the presence of functional motifs using the following software

Table 2.1. Softwares used for the analysis:

Software used for motif search

Programme name	Description	Sequence source	Protein family definition
HMMPfam	Scans the hidden markov models (HMMs) that are present in the Pfam Protein families database.	UniProtKB, GenPept, metagenome data	HMM
ScanProsite	ProfileScan scans against PROSITE profiles. These profiles are based on weight matrices and are more sensitive for the detection of divergent protein families.	UniProtKB/SwissProt	weight matrices
	PatternScan is a new version of the PROSITE pattern search software which uses new code developed by the PROSITE team.	UniProtKB/SwissProt	Regular expressions
FingerPRINTScan	FingerPRINTScan provides the closest matching PRINTS fingerprint(s) to a protein sequence. The latest version of the FingerPRINTScan package uses an approach, based on Karlin and Altschul statistics, to generate a probability-value for each scoring match to a motif.	UniProtKB	protein fingerprints
SMART (Simple Modular Architecture Research Tool)	Scans the hidden markov models (HMMs) that are present in the SMART domain/domain families database.	UniProtKB, ENSEMBL	HMM
HMMPanther	The PANTHER (Protein ANalysisTHrough Evolutionary Relationships) Classification System was designed to classify proteins (and their genes) in order to facilitate high-throughput analysis.	UniProt	HMM
InterProScan Scans sequence using several protein sequence applications.	BlastProDom Scans the families in the ProDom database.	UniProtKB	PSI-BLAST based clusters
	HMMPIR Scans the hidden markov models (HMMs) that are present in the PIR Protein Sequence Database (PSD) of functionally annotated protein sequences, PIR-PSD	microbial sequences from UniProtKB/SwissProt	weight matrices
	HMMTigr Scans the hidden markov models (HMMs) that are present in the TIGRFAMs protein families database	UniProtKB, GenPept, metagenome data	HMM

	HAMAP Scans against HAMAP profiles. These profiles are based on weight matrices and are more sensitive for the detection of divergent bacterial, archaeal and plastid-encoded protein families.	microbial sequences from UniProtKB/SwissProt	weight matrices
	SuperFamily SUPERFAMILY is a library of profile hidden Markov models that represent all proteins of known structure.	UniProtKB, PDB, complete genomes	HMM
	HMMPfam		
	ScanProsite		
	FingerPRINTScan		
	SMART (Simple Modular Architecture Research Tool)		
	HMMPanther		
	TMHMM		
	SignalPHMM		
	Gene3D Gene3D is supplementary to the CATH database. This protein sequence database contains proteins from complete genomes which have been clustered into protein families and annotated with CATH domains, Pfam domains and functional information from KEGG, GO, COG, Affymetrix and STRINGS.	GenBank	HMM

Phylogeny

2.4 Phylgeny.fr

Phylogeny.fr is a free, simple to use web service dedicated to reconstructing and analyzing phylogenetic relationships between molecular sequences. Phylogeny.fr runs and connects various bioinformatics programs to reconstruct a robust phylogenetic tree from a set of sequences.

2.5 Signal peptide and Glycosilation Site

SignalP

SignalP server predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms: Gram-positive prokaryotes, Gram-negative prokaryotes, and eukaryotes. The method incorporates a prediction of cleavage sites and a signal peptide/non-signal peptide prediction based on a combination of several artificial neural networks.

NetNGlyc

The NetNglyc server predicts N-Glycosilation sites in human proteins using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr sequons.

2.6 Molecular Structure

I-TASSER

I-TASSER server is an internet service that generates high-quality predictions of 3D structure and biological function of protein molecules from their amino acid sequences.

- The server first tries to retrieve template proteins of similar folds (or super-secondary structures) from the PDB library by LOMETS, a locally installed meta-threading approach.
- The continuous fragments excised from the PDB templates are reassembled into full-length models by replica-exchange Monte Carlo simulations with the threading unaligned regions (mainly loops) built by ab initiomodeling. In cases where no appropriate template is identified by LOMETS, I-TASSER will build the whole structures by ab initio modeling. The low free-energy states are identified by SPICKER through clustering the simulation decoys.

- The fragment assembly simulation is performed againstarting from the SPICKER cluster centroids, where the spatial restrains collected from both the LOMETS templates and the PDB structures by TM-align are used to guide the simulations. The purpose of the second iteration is to remove the steric clash as well as to refine the global topology of the cluster centroids. The decoys generated in the second simulations are then clustered and the lowest energy structures are selected. The final full-atomic models are obtained by REMO which builds the atomic details from the selected I-TASSER decoys through the optimization of the hydrogen-bonding network.
- For predicting the biological function of the protein, the I-TASSER server matches the predicted 3D models to the proteins in 3 independent libraries which consist of proteins of known enzyme classification (EC) number, gene ontology (GO) vocabulary, and ligand-binding sites. The final results of function predictions are deduced from the consensus of top structural matches with the function scores calculated based on the confidence score of the I-TASSER structural models, the structural similarity between model and templates as evaluated by TM-score, and the sequence identity in the structurally aligned regions.

Results

3. Results of analysis:

3.1 Protein Sequence

The target protein sequence was derived from Genet (UniProtKB/Swiss-Prot ID: P0CE79.1)

>P0CE79.1 RecName: Full=Phospholipase D LrSicTox-alphaIA1ii; Short=PLD; AltName: Full=Dermonecrotic toxin; AltName: Full=SMaseD/LysoPLD; AltName: Full=Sphingomyelin phosphodiesterase D; Short=SMD; Short=SMase D; Short=Sphingomyelinase D; Flags: Precursor

MLLYVTLILGCWSAFSESAETDVAERANKRPIWIMGHMVNAIYQIDEFVNLGAN
SIETDVSFDKDANPEYTYHGVPCDCGRSCLKWEYFSDFLKGLRKATTPGDSKYH
AKLVLVVFDLKTGSLYDNQAYDAGKKLAKNLLKHYWNNGNNGGRAYIVLSIPD
LNHYKLITGFKETLKSEGHPELMDKVGHDFSGNDAIGDVGNAYKKAGVTGHVW
QSDGITNCLLRGLSRVKEAVKNRDSSNGFINKVYYWTVDKRATTREALDAGVD
GVMTNYPDVITDVLNESAYKAKFRIATYDDNPWETFKN

3.2 Homology

PSI BlastPSI-BLAST was used to find homologous sequences for Brown Recluse Phospholipase D in protein databases (GenPept). PSI-BLAST was done up to 3 iteration with max 100 sequence entries. The search was done against non-redundant protein sequence database.

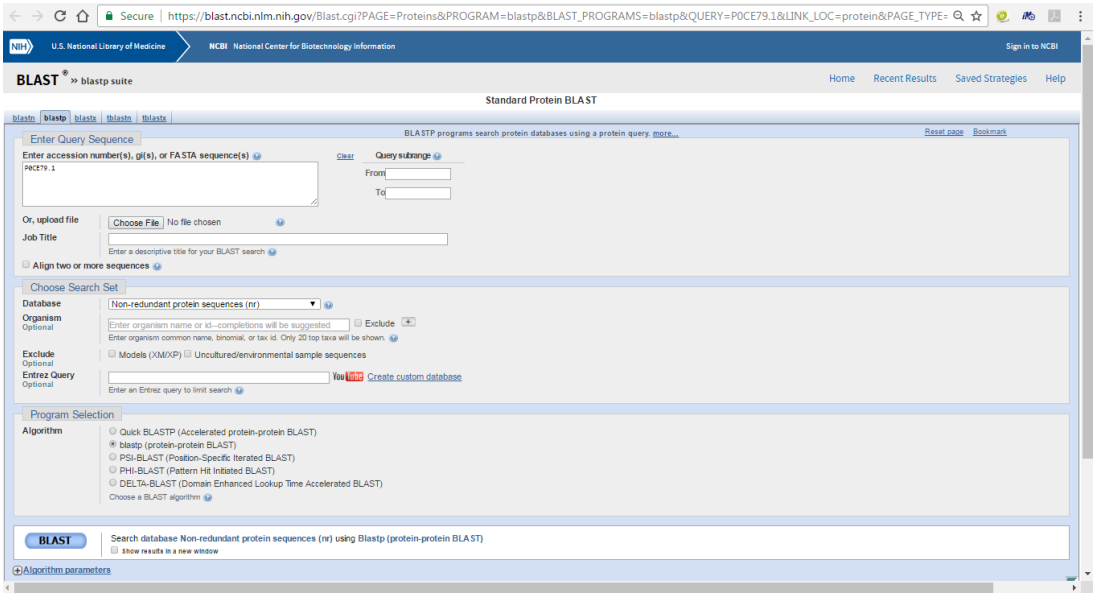

FIG 3.2.1 PSI BLAST Window

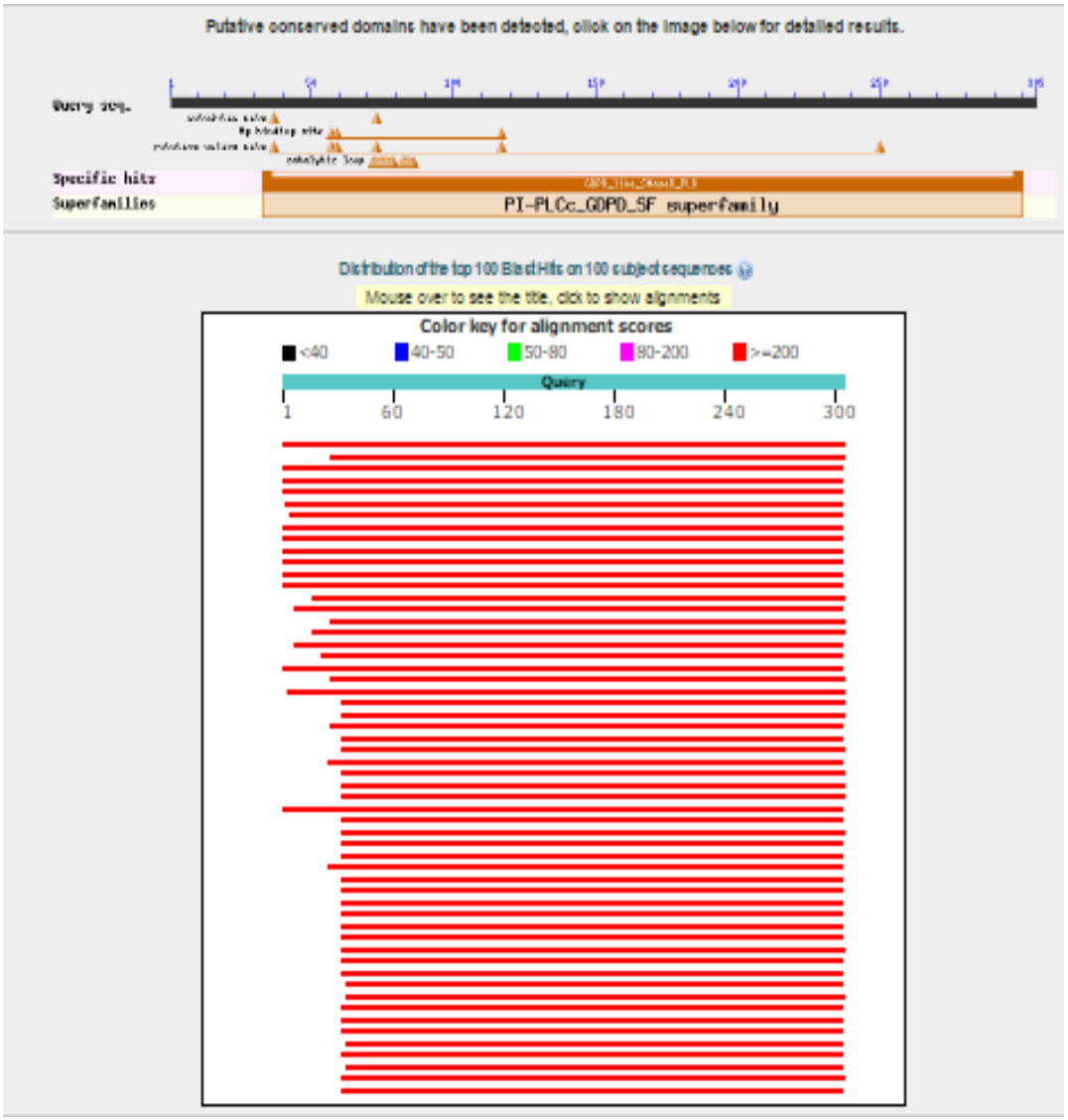

FIG. 3.2.2 PSI-BLAST result

FIG. 3.2.3 PSI-BLAST result

COBALT

Multiple sequence alignment of Brown Recluse Sphingomyelinase D was done with the following 16 spider venom sequences. Alignment type was selected to fast and the protein weight matrix used was BLOSUM. Rest of the programme parameters were default.

AAP97092.2, ANY30963.1, 3RLHA, AAP97091.2, AAQ16123.1, 4RW3A, 3RLGA, 4RW5A, ACN48860.1, AKB91150.1, ACN48838.1, AKB91128.1, AKB91132.1, AAP44735.2, AKB91140.1.

Table 3.2.1 Spider species worked with

Accession Number	Species
P0CE79.1	**Loxosceles reclusa**
AAP97092.2	**Loxosceles intermedia**
ANY30963.1	***Loxosceles similis***
3RLHA	***Loxosceles intermedia***
AAP97091.2	
AAQ16123.1	
4RW3A	
3RLGA	
4RW5A	
ACN48860.1	***Loxosceles spadicea***
AKB91150.1	***Loxosceles rufescens***
ACN48838.1	***Loxosceles hirsuta***
AKB91128.1	***Loxosceles sp. Gran Canaria 1***
AKB91132.1	***Loxosceles sp. La Gomera***
AAP44735.2	***Loxosceles arizonica***
AKB91140.1	***Loxosceles rufescens***

Conserved regions of the sequences were highlighted

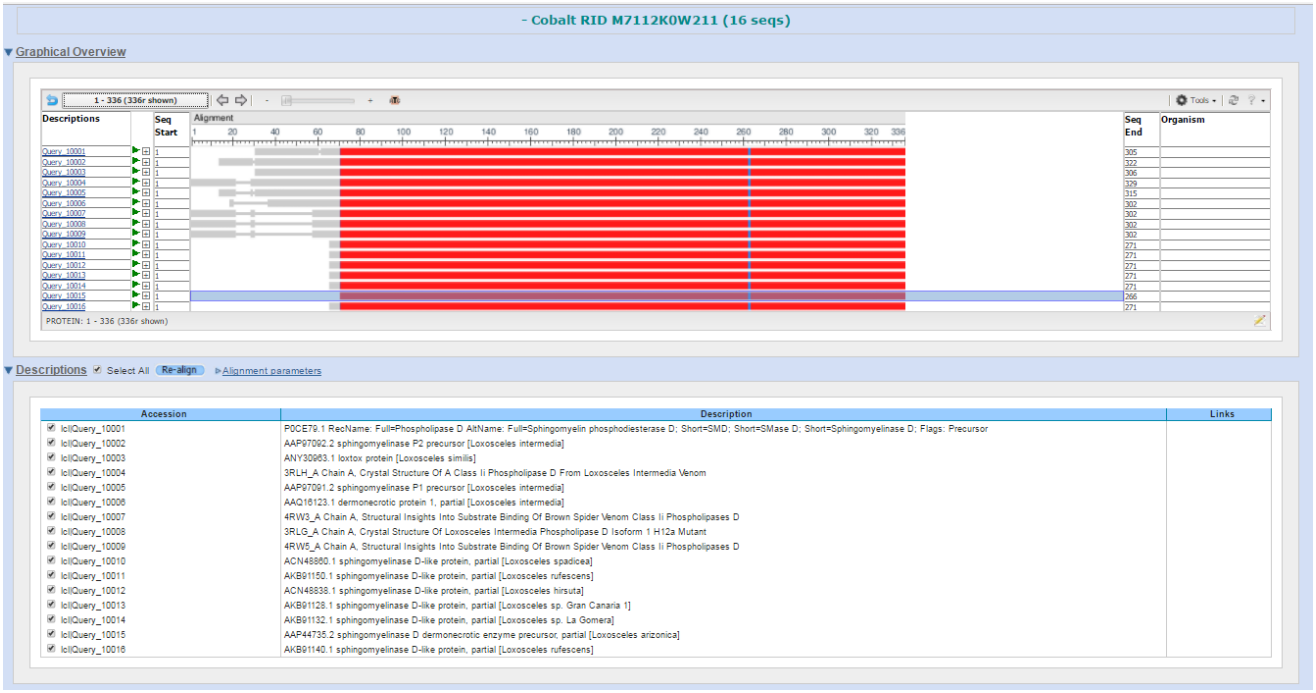

FIG 3.2.4 MSA of the desired protein sequences

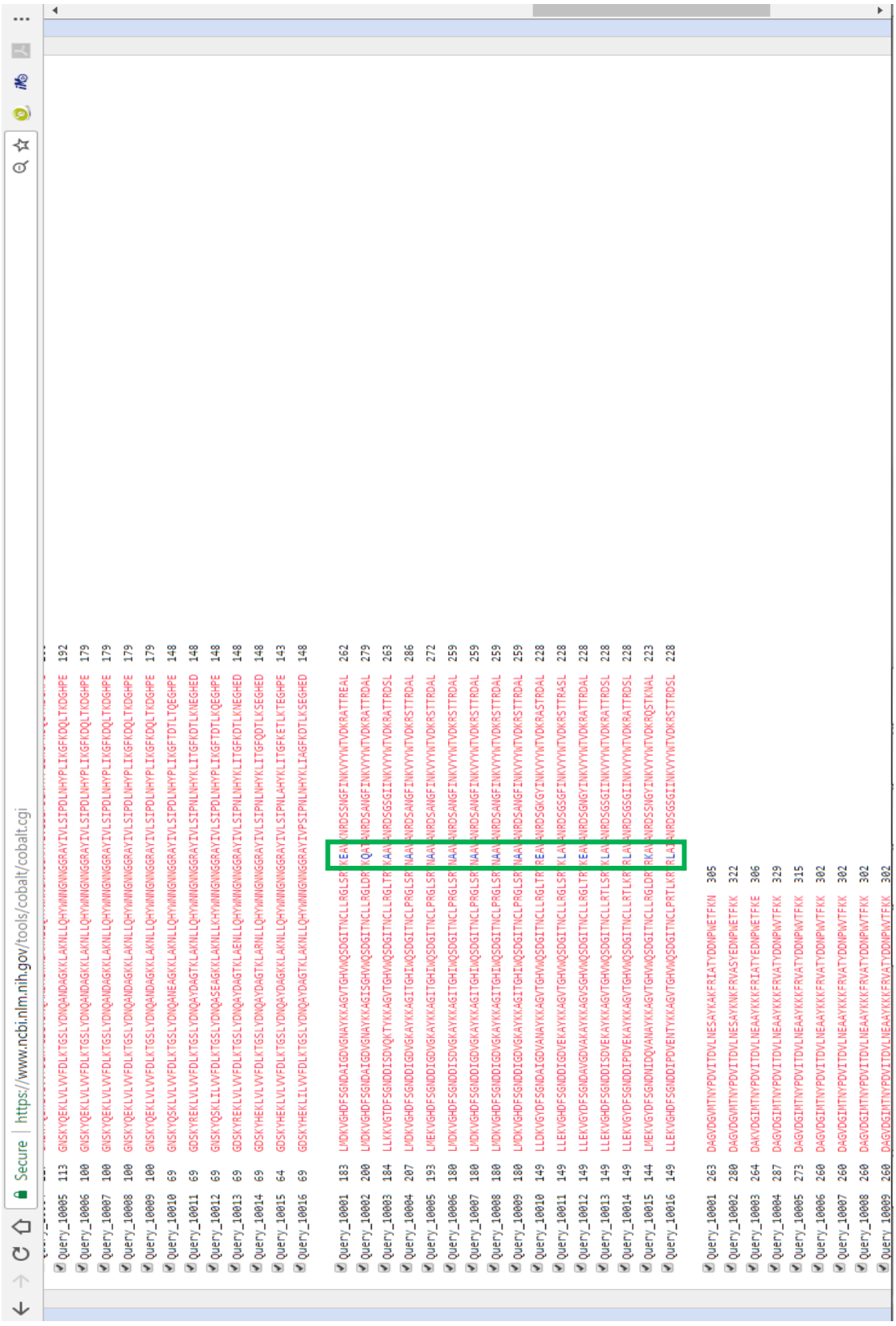

FIG 3.2.5 MSA of Phospholipase D

3.3 Protein Motif

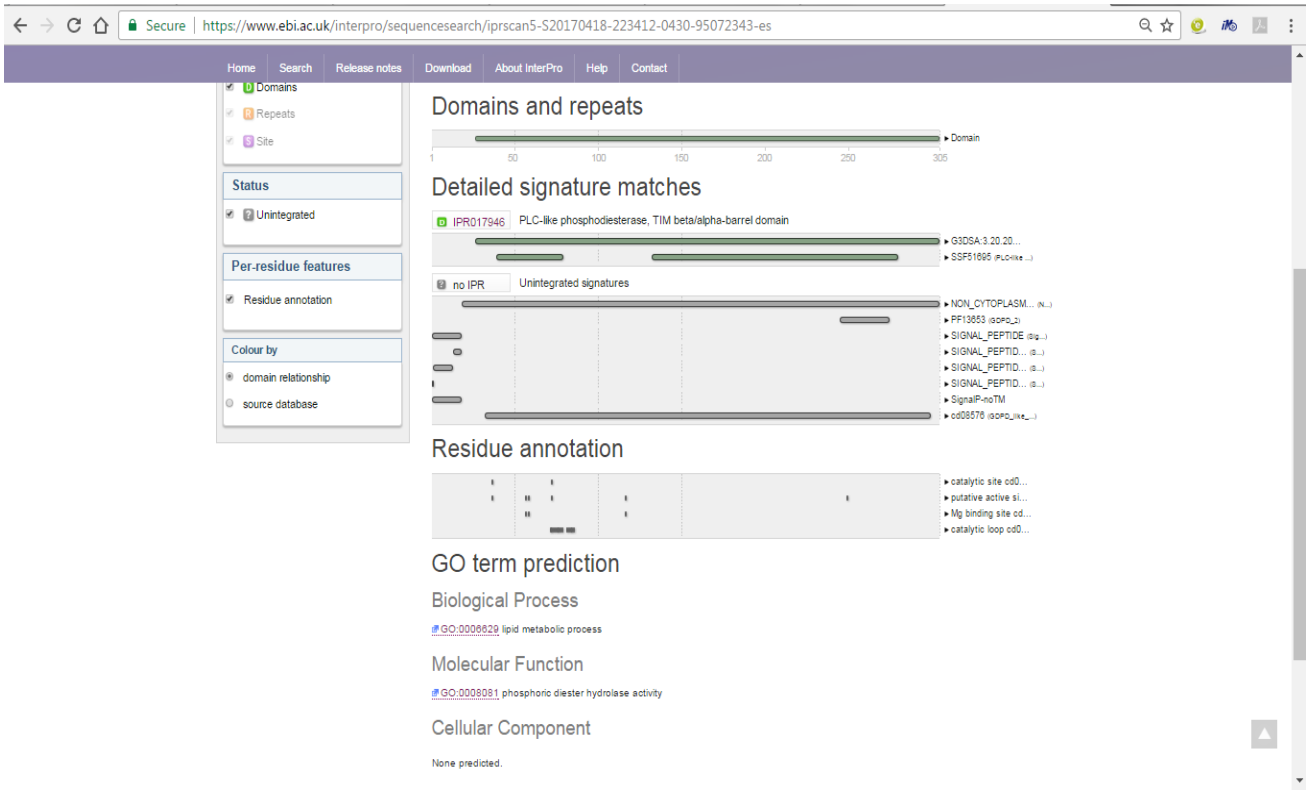

FIG 3.3.1 Predicted Protein Motifs in Phospholipase D

Table 3.2.1 Motifs found in Phospholipase D in Brown Recluse

Software's/ Database	Motifs/ Domains	Position in Sequence	E-Value
Pfam	GDPD 2	246-275	0.0039
Superfamily	PLC- like Phosphodiesterases	40-79	6.26e-17
	Glycerophosphoryl diester diesterase	133-280	0.018
Signal1P	Signal Peptide	1-18	0.768

3.4 Phylogeny

To determine the phylogeny of Brown Recluse Phospholipase D, total 16 protein sequences were taken.

Phospholipase D sequence of the following spider species- *Loxeceles Recluse (P0CE79.1), Loxosceles Intermedia (AAP97092.2), Loxosceles Similis (ANY30963.1), Loxosceles Spadicea (ACN48860.1), Loxosceles Rufescens (AKB91150.1), Loxosceles Hirsuta (ACN48838.1), Loxosceles sp. Gran Canaria (AKB91128.1), Loxosceles sp. La Gomera (AKB91132.1), Loxosceles arizonica (AAP44735.2), Loxosceles Intermedia (3RLHA), Loxosceles Intermedia (AAP97091.2), Loxosceles Intermedia (AAQ16123.1), Loxosceles Intermedia (4RW3A), Loxosceles Intermedia (3RLGA), Loxosceles Intermedia (4RW5A)*

Table 3.4.1: E-value, Max Identity, and Max Score of different *Loxoceles* Species

Species	Protein	E value	Max identity	Max score
Loxeceles Recluse	PLD	0	100%	633
Loxoceles Intermedia	PLD	0	85%	553
Loxoceles Similis	PLD	0	84%	543
Loxoceles Spedicea	PLD	3e-176	87%	486
Loxoceles Rufescens	PLD	6e-176	87%	496
Loxoceles Hirsuta	PLD	4e-175	86%	494
Loxoceles Gran Canaria	PLD	1e-172	86%	487
Loxoceles Intermedia	PLD	0	83%	538
Loxosceles Intermedia	PLD	0	83%	537

Species	Protein	E value	Max identity	Max score
Loxosceles Intermedia	PLD	0	83%	528
Loxosceles Intermedia	PLD	4e-179	85%	505
Loxosceles Intermedia	PLD	7e-178	85%	501

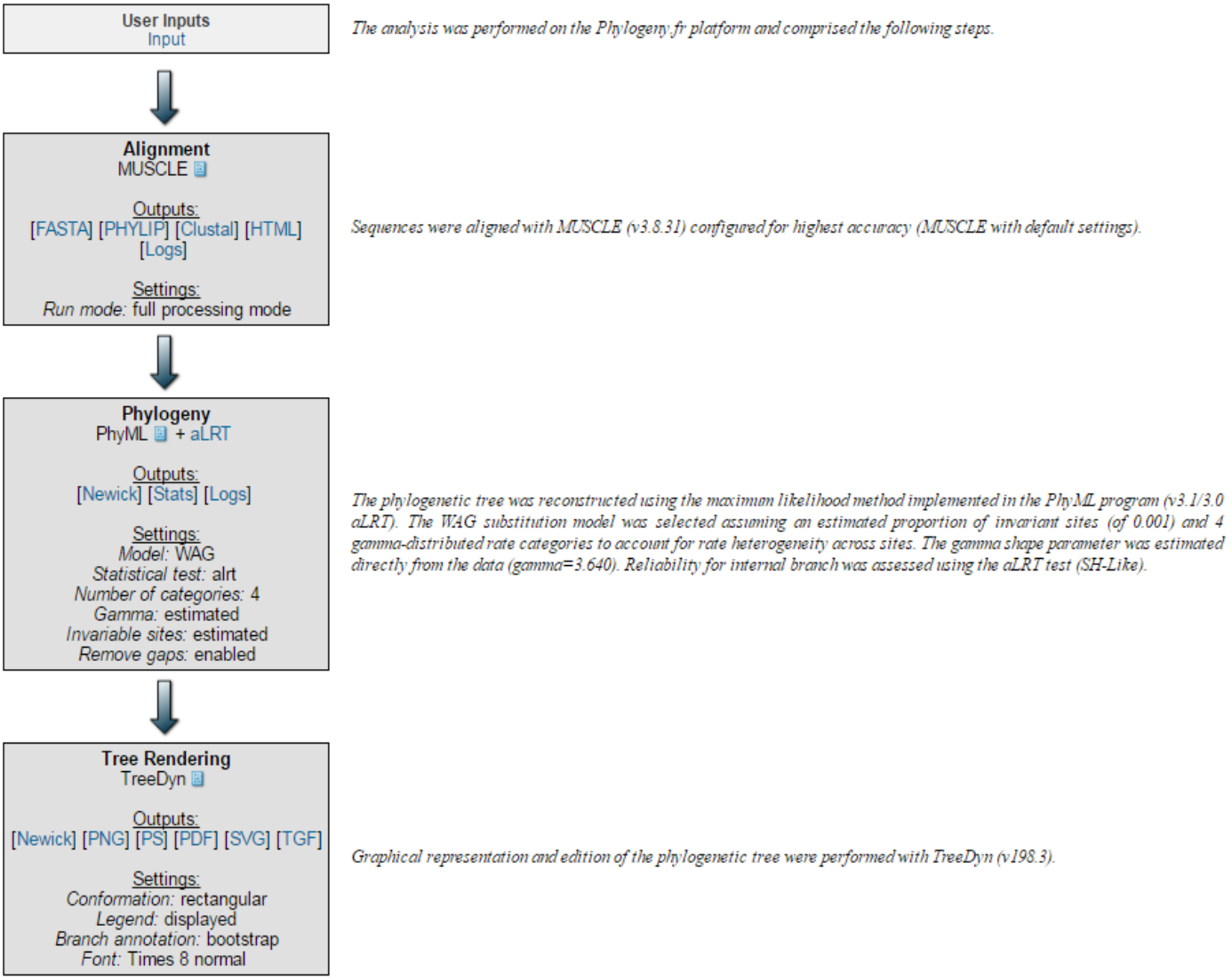

FIG 3.4.1 Flow Chart of phylogeny

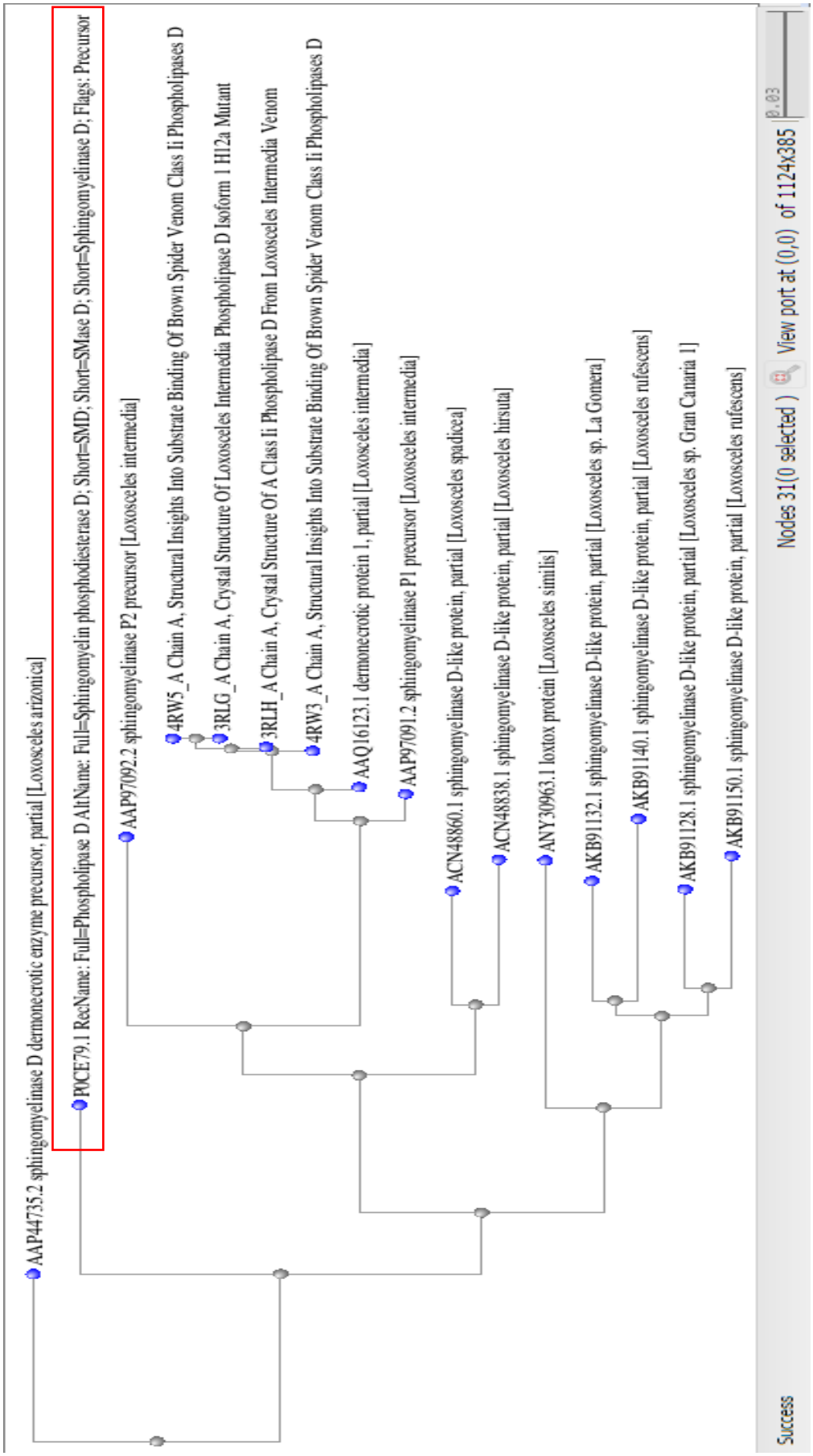

FIG 3.4.2 Phylogenic Tree

3.5 Signal Peptide and Glycosilation site

Signal P

FIG 3.5.1 the result of SignalP

SignalP (version 4.1) predicted the presence of a 20 a. a. signal peptide for the Brown Recluse Phospholipase D. The probable cleavage site is between Asn[18] and His[19].

C-score (raw cleavage site score) -The output from the CS networks, which are trained to distinguish signal peptide cleavage sites from everything else. S-score (signal peptide score) - The output from the SP networks, which are trained to distinguish positions within signal peptides from positions in the mature part of the proteins and from proteins without signal peptides. Y-score (combined cleavage site score) - A combination (geometric average) of the C-score and the slope of the S-score, resulting in a better cleavage site prediction than the raw C-score alone.

NetNGlyc

FIG 3.5.2 Predicted N- Glycosilation site

NetNglyc server predicted that there was no glycosilation site in the target protein.

3.6 Molecular Structure:

I-TASSER server returned the following results for Brown Recluse spider venom Phospholipase D:

Submitted sequence

>protein
MLLYVTLILGCWSAFSESAETDVAERANKRPIWIMGHMVNAIYQIDEFVNLGAN
SIETDVSFDKDANPEYTYHGVPCDCGRSCLKWEYFSDFLKGLRKATTPGDSKYH
AKLVLVVFDLKTGSLYDNQAYDAGKKLAKNLLKHYWNNGNNGGRAYIVLSIPD
LNHYKLITGFKETLKSEGHPELMDKVGHDFSGNDAIGDVGNAYKKAGVTGHVW
QSDGITNCLLRGLSRVKEAVKNRDSSNGFINKVYYWTVDKRATTREALDAGVD
GVMTNYPDVITDVLNESAYKAKFRIATYDDNPWETFKN

FIG 3.6.1 Submitted protein sequence to I-TASSER Server

Predicted Secondary Structure

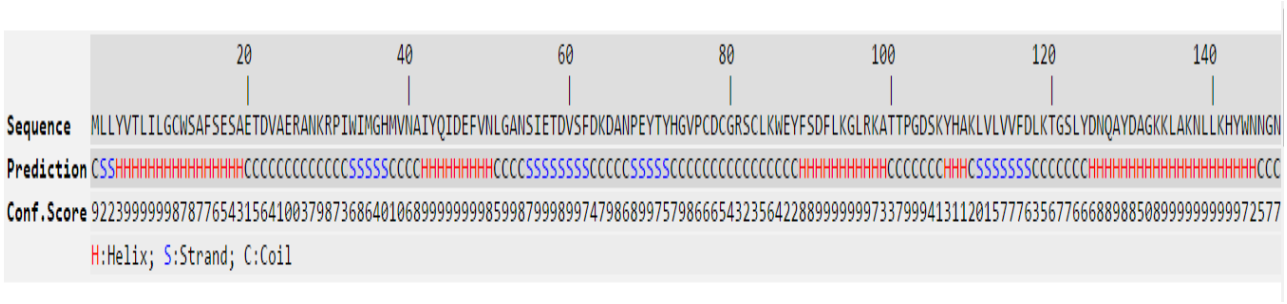

FIG 3.6.2 Predicted secondary structure of Brown Recluse Phospholipase D

Predicted Solvent Accessibility

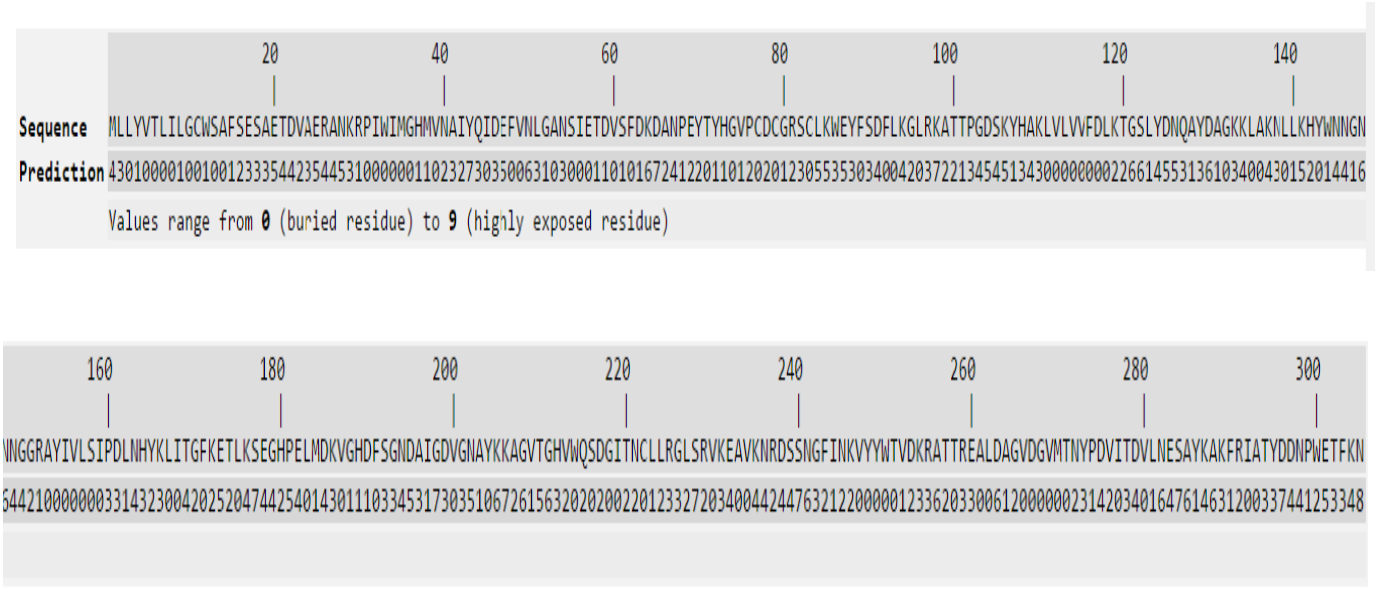

FIG 3.6.3 Predicted solvent accessibility of Brown Recluse Phospholipase D

Top 10 templates used by I-TASSER

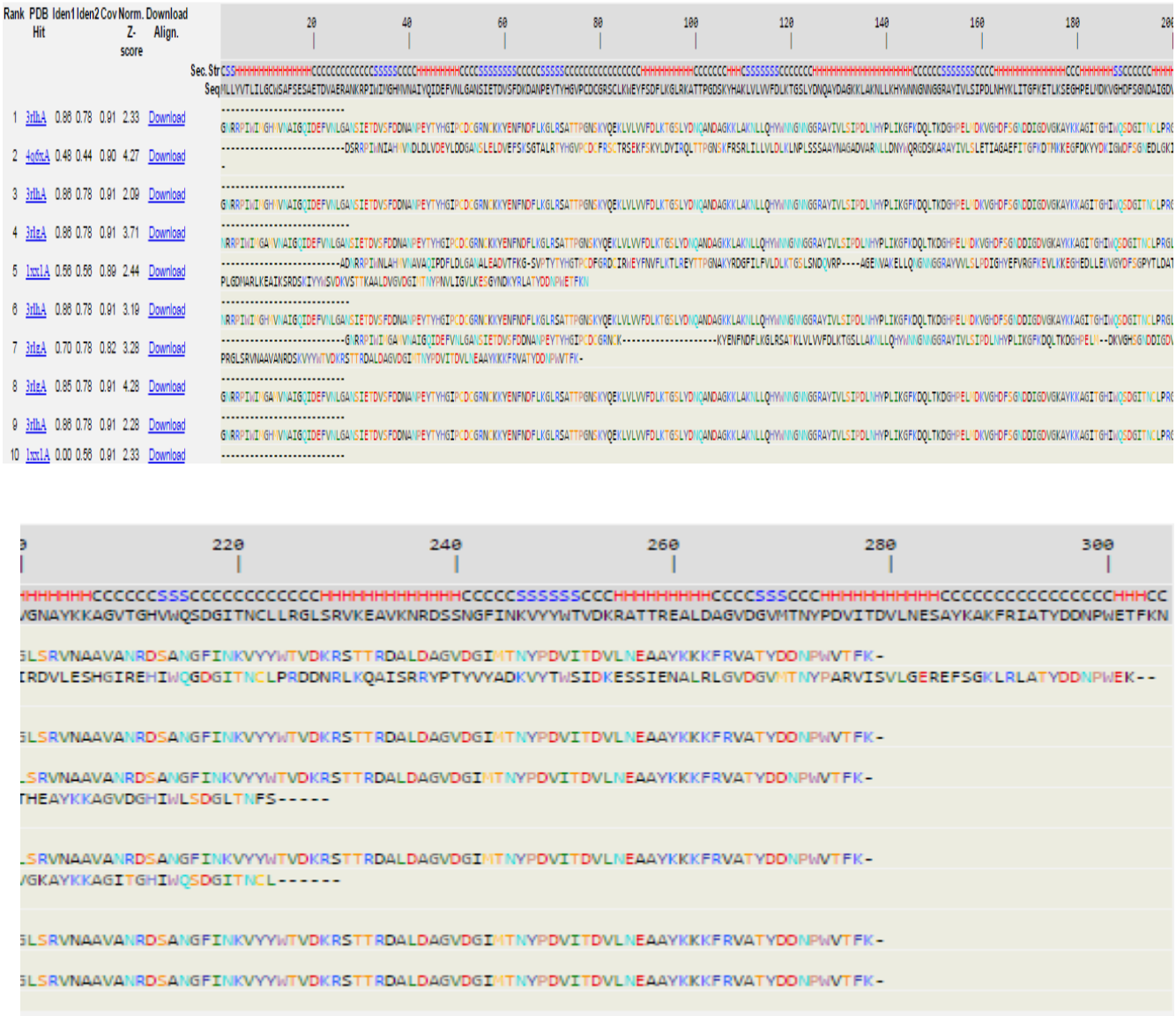

FIG 3.6.4 Alignment with top 10 protein templates used by I-TASSER

The data provided that-

All the residues are colored in black; however, those residues in template which are identical to the residue in the query sequence are highlighted in color. Coloring scheme is based on the property of amino acids, where polar are brightly colored while non-polar residues are colored in dark shade.

- Rank of templates represents the top ten threading templates used by I-TASSER.
- Ident1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.
- Ident2 is the percentage sequence identity of the whole template chains with query sequence.
- Cov represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein.
- Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.
- Download Align. Provides the 3D structure of the aligned regions of the threading templates.
- The top 10 alignments reported above are from the following threading programs:

1: MUSTER 2: FFAS-3D 3: SPARKS-X 4: HHSEARCH2 5: HHSEARCH I 6: Neff-PPAS 7: HHSEARCH 8: pGenTHREADER 9: wdPPAS 10: cdPPAS

The top 5 models selected by I-TASSER are as follows:

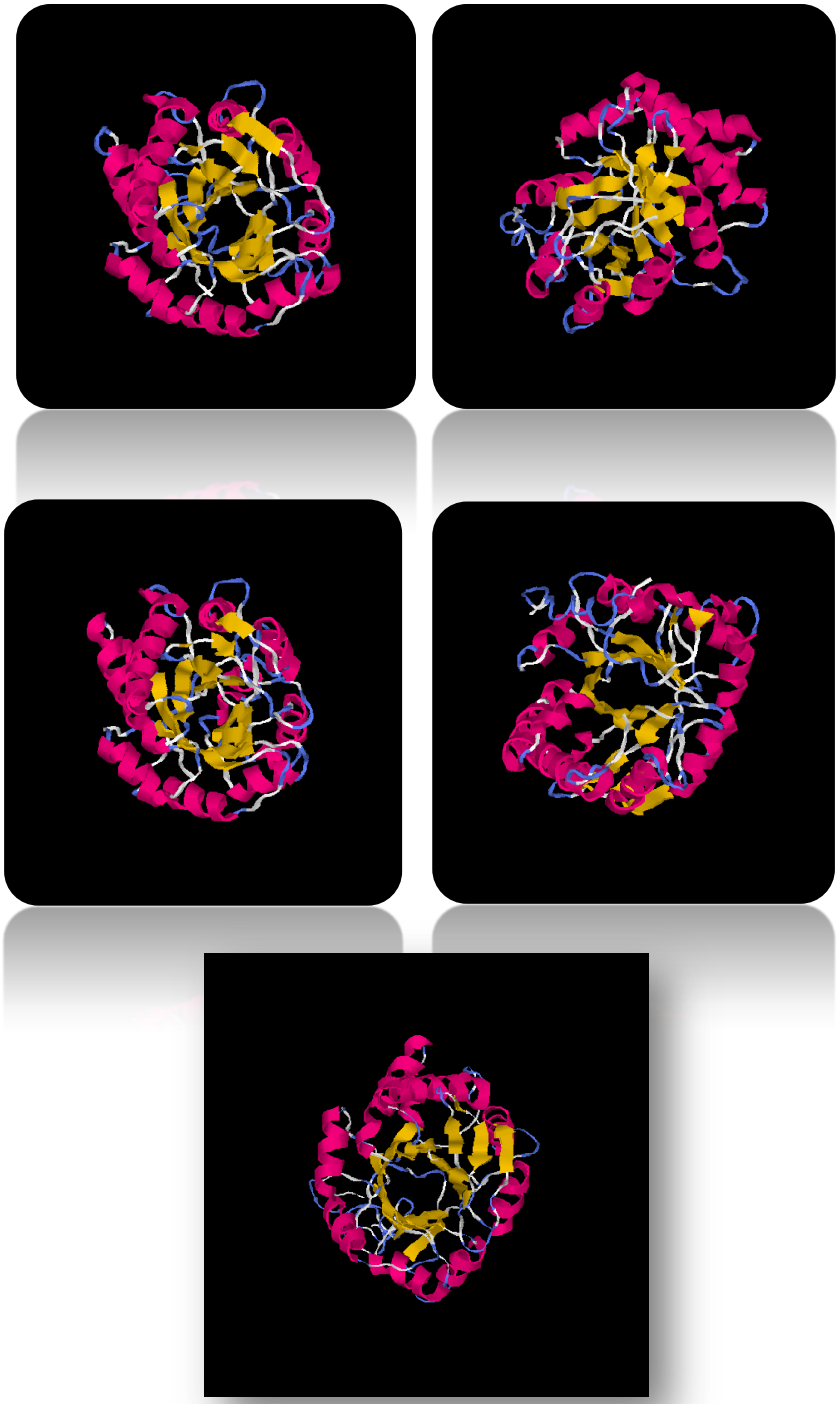

Figure 3.6.5: Top 2 (from left to right, rank 1-2) predicted models with C-scores -0.51 and 1.49 respectively. 2 models in 2nd row rank 3-4 (from left to right) with C-scores respectively -1.24, - 0.90. Model on lower row (rank 5) with C-score -2.81. Estimated accuracy of Model 1: 0.65±0.13 (TM-score), 7.3±4.2Å (RMSD). C-score is a confidence score for estimating the quality of predicted models by I-TASSER

TABLE 3.6.1

Top 10 Identified structural analogs in PDB

Rank	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov
1	3rlgA	0.906	0.58	0.853	0.911
2	4q6xA	0.887	0.95	0.486	0.905
3	2f9rD	0.887	1.37	0.612	0.911
4	2vx7A	0.644	4.26	0.068	0.862
5	4yn5A	0.640	4.42	0.059	0.869
6	2bvyA	0.639	4.36	0.071	0.859
7	4zxoA	0.638	4.17	0.076	0.839
8	3i6rA	0.636	4.24	0.094	0.830
9	4lypA	0.634	4.55	0.067	0.862
10	3cuzA	0.632	4.40	0.081	0.820

- Query structure is shown in cartoon, while the structural analog is displayed using backbone trace.
- Ranking of proteins is based on TM-score of the structural alignment between the query structure and known structures in the PDB library.
- RMSDa is the RMSD between residues that are structurally aligned by TM-align.
- IDENa is the percentage sequence identity in the structurally aligned region.
- Cov represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

TABLE 3.6.2

Top Ligand Binding site

Rank	C-score	Cluster size	PDB Hit	Lig Name	Ligand Binding Site Residues
1	0.27	13	3qvqB	G3P	37,38,57,59,73,117,119,160,192,250,270
2	0.10	8	3rlgA	MG	57,59,117
3	0.03	4	1djwA	CIP	37,38,40,57,59,73,117,269,271
4	0.03	4	1GYMA	1GYMA00	160,192,250
5	0.02	3	2OOGD	2OOGD05	54,109,110,111,112,152,294,295

- C-score is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a more reliable prediction.
- Cluster size is the total number of templates in a cluster.
- Lig Name is name of possible binding ligand. Click the name to view its information in the BioLiP database.
- Rep is a single complex structure with the most representative ligand in the cluster, i.e., the one listed in the Lig Name column.
- Mult is the complex structures with all potential binding ligands in the cluster.

TABLE 3.6.3

Top5 5 enzyme homologous in PDB

Rank	$Cscore^{EC}$	PDB Hit	TM-score	$RMSD^a$	$IDEN^a$	Cov	EC Number	Active Site Residues
1	0.579	1xx1A	0.888	1.56	0.609	0.915	3.1.4.41	37,59,73,74,78,119,250,253,272
2	0.518	2f9rA	0.886	1.44	0.612	0.911	3.1.4.41	268
3	0.300	1vlwB	0.510	3.19	0.138	0.610	4.1.3.16	57
4	0.235	3i68A	0.634	4.32	0.097	0.836	1.3.5.2	NA
5	0.234	2gq3A	0.624	4.47	0.091	0.820	2.3.3.9	NA

- $Cscore^{EC}$ is the confidence score for the EC number prediction. CscoreEC values range in between [0-1]; where a higher score indicates a more reliable EC number prediction.

- TM-score is a measure of global structural similarity between query and template protein.
- RMSDa is the RMSD between residues that are structurally aligned by TM-align.
- IDENa is the percentage sequence identity in the structurally aligned region.
- Cov represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

TABLE 3.6.4

Predicted GO (Gene ontology) terms

Rank	Cscore^{GO}	TM-score	RMSD^{a}	IDEN^{a}	Cov	PDB Hit	Associated GO Terms
1	0.34	0.8877	1.56	0.61	0.91	1xx1A	GO:0019835 GO:0044179 GO:0005576 GO:0016787 GO:0046872 GO:0050290 GO:0006629 GO:0008081
2	0.24	0.6414	4.26	0.06	0.86	2vx4A	GO:0003824 GO:0005975 GO:0006080 GO:0008810 GO:0016985 GO:0043169
3	0.24	0.6339	4.32	0.10	0.84	3i68A	GO:0003824 GO:0004152 GO:0006207 GO:0006222 GO:0008152 GO:0055114 GO:0004158 GO:0016020
4	0.24	0.6382	4.37	0.07	0.86	2bvtA	GO:0003824 GO:0005975 GO:0006080 GO:0008810 GO:0016985 GO:0043169
5	0.23	0.6244	4.47	0.09	0.82	2gq3A	GO:0005737 GO:0006099 GO:0015936 GO:0004474 GO:0030145 GO:0016740 GO:0005618 GO:0005886 GO:0005829 GO:0000287 GO:0005576 GO:0003824
6	0.23	0.6321	4.40	0.08	0.82	3cuzA	GO:0004474 GO:0016740 GO:0005737 GO:0005515 GO:0006099 GO:0006097 GO:0003824
7	0.23	0.6304	4.59	0.05	0.86	2qhaA	GO:0046872 GO:0016798 GO:0016787 GO:0016985 GO:0008152 GO:0003824 GO:0005975 GO:0006080 GO:0008810 GO:0005576
8	0.22	0.6055	2.93	0.18	0.69	3no3A	GO:0046872 GO:0006071 GO:0006629 GO:0008081 GO:0008889
9	0.21	0.5396	3.28	0.12	0.66	3exrA	GO:0008152 GO:0004590 GO:0003824 GO:0006207
10	0.21	0.5582	2.67	0.18	0.63	1vd6A	GO:0008081 GO:0006071 GO:0035556 GO:0004629 GO:0005622 GO:0006629 GO:0008889

TABLE 3.6.5

Consensus Prediction of Gene Ontology terms

Molecular Function	GO:0016634	GO:0016635	GO:0016985	GO:0008810	GO:0046914	GO:0050290				
GO-Score	0.47	0.47	0.42	0.42	0.36	0.34				
Biological Process	GO:0019856	GO:0009174	GO:0046049	GO:0046487	GO:0033865	GO:0046356	GO:0045333	GO:0046128	GO:0006080	GO:0019835
GO-Score	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.42	0.34
Cellular Component	GO:0005576	GO:0044444	GO:0030312							
GO-Score	0.49	0.47	0.47							

- CscoreGO is a combined measure for evaluating global and local similarity between query and template protein. Its range is [0-1] and higher values indicate more confident predictions.
- TM-score is a measure of global structural similarity between query and template protein.

- RMSDa is the RMSD between residues that are structurally aligned by TM-align.
- IDENa is the percentage sequence identity in the structurally aligned region.
- Cov represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.
- The second table shows a consensus GO terms amongst the top scoring templates. The GO-Score associated with each prediction is defined as the average weight of the GO term, where the weights are assigned based on CscoreGO of the template.

Discussion

4. Discussion:

Based upon the results obtained from this study, the following can be inferred

- The target protein sequence of the desired spider species is very much conserved within the spider species (*Loxoceles* and other genuses) only. There have been some evolutionary changes from time to time in the past years but still Phospholipase D from each species is very much similar to each other.
- No significant similarity was found to proteins with similar functions in other insect or animal species.
- The result of the Multiple Sequence Alignment using COBALT shows that there is a very long stretch of conserved region in the whole sequence.
- For the MSA the species having the maximum E value has been taken. This E value is actually the confidence value that is how much similarities are there with the protein sequences found and the desired protein.
- Only one single amino acid in region which is in the range of 50 was found to be different in all the sequences selected for MSA. Similar kind of result was predicted in case of CHH hormone in black widow [19]. It will be interesting to explore whether this single amino acid is responsible for substrate binding preferences or other attributes that make the target protein slightly different from other Phospholipase D.
- The predicted signal peptide is a 20 a.a.long with cleavage between Asn^{18} and His^{19}.
- The only functional motif spanning the whole protein is PLC- like Phosphodiesterases, based on the results from several protein motif detection software.
- The protein motifs were found with the software Interpro Scan. It predicted that there is possible cleavage site. The motif configuration of CHHs differs from that of molt-inhibiting hormone or gonad-inhibiting hormone in both N and C termini. These two motifs approach each other in tertiary structure models, and the motif preference reveals the critical roles of these regions in functional specificity [19].
- The protein motif finding software's were able to find some motif or binding sites. Namely three binding sites were seen, GDPD, PLC and Signal peptides. Glycerophosphodiester Phosphodiesterases (GDPD) is enzyme which degrades

various glycerophosphodiesters to produce glycerol-3-phosphate and the corresponding alcohol moiety. Apart from this, a very interesting finding is that this enzyme could be used in the degradation of toxic organophosphorus esters, which has resulted in much attention on the biochemical and application research of GDPDs [21]. Phospholipase C is Phosphodiesterases. PLC is a family of eukaryotic intracellular enzymes that play an important role in signal transduction processes [22]. Signal peptides are a cornerstone mechanism for cellular protein localization, yet until now experimental determination of signal peptides has come from only a narrow taxonomic sampling. As a result, the dominant view is that Sec-cleaved signal peptides in prokaryotes are defined by a canonical AxA motif. Although other residues are permitted in the motif, alanine is by far the most common [23]

- The NetNGlyc software predicted that there is no possible glycosilation sites in the target protein.
- Upon phylogenetic analysis, the Phospholipase D from the Brown Recluse forms a separate group of its own very much distinct from the other spider species. This finding can be used for future works.
- In the phylogenic tree created we observe that our desired protein sequence has formed a group alone and the others are present in the sub groups. This also proves that the protein sequence has some similarities with other species but then again highly conserved among them. To make a comparison the results of the same spider of a different species were seen. Phylogenetic tree analysis revealed a structural relationship for these toxins compared to other members of family. Recombinant molecules were expressed as N-terminal His-tag fusion proteins in *Escherichia coli* and were purified to homogeneity from cell lysates by Ni^{2+} chelating chromatography, resulting in proteins of 33.8 kDa for LiRecDT2 and 34.0 kDa for LiRecDT3 [20].
- The results obtained from the I-TASSER server predict that in the protein structure, the alpha helices form the outer coat whereas the beta sheets form a core in the middle and the strands help them to connect with each other. Most of the conserved residues are in the putative FAD binding region. It can be hypothesized that the core region is involved in substrate binding and enzymatic function while the alpha helices region might have regulatory functions.

The structural and functional analysis of Phospholipase D from the Brown Recluse spider generated some interesting results. Further analysis of the structure of Phospholipase D to elucidate the substrate binding site of the enzyme could help in understanding how the enzyme functions in vivo. Detailed study into the substrate the enzyme works on and the metabolic pathways ultimately affected by its function could clarify our understanding of the big picture. This would translate into development of potential therapeutic lead molecules from the enzyme protein in future.

References

1. Role of spiders in agriculture and horticulture ecosystem, Rajeswaran J, Duraimurugan P, Shanmugam P, 2005, issue: 4, pages: 147-152
2. Checklist of Spiders (Arachnida: Araneae) of South Asia including the 2006 update of Indian Spider Checklist, Siliwal M, Molur Sanjay, Zoos' Print, 2006 volume: 22 (2) pp: 2551-2597
3. Spiders – the Creepy Crawlers We Need | Bird-X Blog, Jean Burr, 2015
4. Loxosceles Reclusa (Brown Recluse) (Non-Prophet), Nature, 2006
5. http://www.wikiwand.com/en/Brown_recluse_spider
6. Brown Recluse Spiders: Facts, Bites & Symptoms, Jessie Szalay, Live Science Contributor, November, 2014
7. Cameron, H. D. 2005. Chapter 73 — An etymological dictionary of North American spider genus names, page 73 in D. Ubick, P. Paquin, P.E. Cushing, and V. Roth (eds.) Spiders of North America: an identification manual. American Arachnological Society, Keene (New Hampshire).
8. Sandidge, Jamal S. and Jennifer L. Hopwood. 2005. Brown recluse spiders: a review of biology, life history and pest management. Transactions of the Kansas Academy of Science 108 p. 99-108.
9. NIH, National Institute of General Medical Sciences (NIGMS). (2016, October 28). Exploring the evolution of spider venom to improve human health. ScienceDaily. Retrieved May , 2017
10. Rash, L.D. and Hodgson, W.C. Pharmacology and biochemistry of spider venoms. Toxicon, 2002, 40 (3) 225-254.
11. Odell, George V., Clement, Herlinda, Possani, Lourival, & Alagón, Alejandro. (2001). Spider Venom Toxins. *Journal of Venomous Animals and Toxins*, *7*(2), 335.
12. Chaim, O. M., Trevisan-Silva, D., Chaves-Moreira, D., Wille, A. C. M., Ferrer, V. P., Matsubara, F. H. Veiga, S. S. (2011). Brown Spider (*Loxosceles* genus) Venom Toxins: Tools for Biological Purposes. *Toxins*, *3*(3), 309–344.
13. Wolfgang Nentwig, Lucia Kuhn-Nentwig, Main Components of Spider Venoms, Spider Ecophysiology, 2013, pp 191-202
14. Spider Bite: Brown Recluse Spider Bite Jerry R. Balentine, DO, FACEP, 2016
15. Bruntz, R. C., Lindsley, C. W., & Brown, H. A. (2014). Phospholipase D Signaling Pathways and Phosphatidic Acid as Therapeutic Targets in Cancer. Pharmacological Reviews, 66(4), 1033–1079.

16. Lajoie DM, Zobel-Thropp PA, Kumirov VK, Bandarian V, Binford GJ, Cordes MHJ (2013) Phospholipase D Toxins of Brown Spider Venom Convert Lysophosphatidylcholine and Sphingomyelin to Cyclic Phosphates. PLoS ONE 8(8): e72372.
17. Phospholipase D from Loxosceles laeta Spider Venom Induces IL-6, IL-8, CXCL1/GRO-α, and CCL2/MCP-1 Production in Human Skin Fibroblasts and Stimulates Monocytes Migration, Rojas J, Arán-Sekul T, Cortés E, Jaldín R, Ordenes K, 2017, vol: 9 (4) pp: 125
18. Analysis of L-amino acid oxidase from King cobra venom and Importance of Snake Database, Abira Khan, 2012
19. JOUR, Chen, S.H.,Lin, C.Y., Kuo, C.M., In Silico Analysis of Crustacean Hyperglycemic Hormone Family, Marine Biotechnology. 2005, Pages:193-206, VL - 7,IS - 3,SN - 1436-2236,DO - 10.1007/s10126-004-0020-5,ID - Chen2005
20. Molecular cloning and functional characterization of two isoforms of dermonecrotic toxin from Loxosceles intermedia (Brown spider) venom gland, Bertoni da Silveira R, Pigozzo R, Chaim O, Appel M, Dreyfuss J, **Biochimie,** 2006 vol: 88 (9) pp: 1241-1253
21. Wang, F., Lai, L., Liu, Y., Yang, B., & Wang, Y. (2016). Expression and Characterization of a Novel Glycerophosphodiester Phosphodiesterase from Pyrococcus furiosus DSM 3638 That Possesses Lysophospholipase D Activity. International Journal of Molecular Sciences, 17(6), 831. http://doi.org/10.3390/ijms17060831
22. Meldrum E, Parker PJ, Carozzi A (1991). "The PtdIns-PLC superfamily and signal transduction". Biochim. Biophys. Acta. 1092 (1): 49–71. PMID 1849017. doi:10.1016/0167-4889(91)90177-Y.
23. Payne SH, et al. 2012. Unexpected diversity of signal peptides in prokaryotes. mBio 3(6):e00339-12. doi:10.1128/mBio.00339-12.

Appendices

ABPPENDICES

BROWSER NAME	WEB ADDRESS
National Centre for Biotechnology Information	https://www.ncbi.nlm.nih.gov/
BLAST	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Interproscan	http://www.ebi.ac.uk/Tools/pfa/iprscan/
COBALT	https://www.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi
Gene Ontology	http://www.geneontology.org/
SignalP	http://www.cbs.dtu.dk/services/SignalP/
NetNGlyc	http://www.cbs.dtu.dk/services/NetNGlyc/
I-TASSER	http://zhanglab.ccmb.med.umich.edu/I-TASSER/
Pfam	http://pfam.xfam.org/