Journal of Pharmacognosy and Phytochemistry

Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
JPP 2018; 7(1): 290-295
Received: 21-11-2017
Accepted: 23-12-2017

Mst. Ferdowsi Mahal
Associate Professor, Department of Botany, University of Rajshahi, Rajshahi, Bangladesh

Md. Azizul Islam
State Key Laboratory of Plant Genomics, Institute of Microbiology, University of Chinese Academy of Sciences, China

Md. Kazy Ebnul Hasan
Department of Genetic Engineering and Biotechnology University of Rajshahi, Rajshahi, Bangladesh

Razia Sultana
Department of Genetic Engineering and Biotechnology, University of Rajshahi, Rajshahi, Bangladesh

Md. Samiul Islam
Department of Genetic Engineering and Biotechnology, University of Rajshahi, Rajshahi, Bangladesh

Md. Amit Hasan
Department of Genetic Engineering and Biotechnology University of Rajshahi, Rajshahi, Bangladesh

Correspondence
Mst. Ferdowsi Mahal
Associate Professor, Department of Botany, University of Rajshahi, Rajshahi, Bangladesh

A computational characterization and motif election analysis of metalloproteinase proteins from different venomous snake species

Mst. Ferdowsi Mahal, Md. Azizul Islam, Md. Kazy Ebnul Hasan, Razia Sultana, Md. Samiul Islam and Md. Amit Hasan

Abstract
Snake venom is a complex mixture of many proteins and polypeptides which have pharmacological activities. These protein components usually comprise a highly evolved and organized armory of biomolecules that attack, often with precision and coordinated synergism, a wide range of molecular targets vital for physiological processes. Snake venom metalloproteinases (SVMPs) are one of the main toxin proteins widely distributed in snake venoms and play significant roles in neurodegenerative disorders, hemostatic disorders and local tissue damage that follows snakebite. The impact of SVMPs on hemostasis has been extensively studied showing diverse effects on soluble factors as well as cellular components. The action of SVMPs involves catalytic and anti-adhesive properties with direct cellular activation or the release of endogenous bioactive components. In this study, a total of 16 metalloproteinases protein sequences were retrieved from NCBI protein database and characterized for different physic-chemical properties, Multiple Sequence Alignment (MSA), phylogenetic analysis and motifs election. ExPASy's Prot Param tool was used to analyze the physic-chemical properties of the selected proteins. The Molecular Weight (M.Wt) ranges from 27813.6 to 69554.9 Da, Isoelectric points (pI) of ten distinct snake species were found to be acidic in nature and remaining six basic behaviors. The aliphatic index figure outs studied metalloproteinases illustrated the tendency of having both wide and low range of temperature as 9 proteins showed AI above 70 and others showed AI value below 70. The lower GRAVY value signifies that there will be better interaction of protein with water. The secondary structure prediction was carried out by SOPMA which showed that random coils dominated all the other conformations. Multiple sequence analysis and Phylogenetic analysis of metalloproteinases were done by MEGA 5. Motif election was completed by MEME which signifies motif 1 (13 sequences), Motif 2 (15 sequences), Motif 3 (15 sequences) and these shows the region which was done by Pfam also allude the reprolysin protein family.

Keywords: snake venom, metalloproteinase, bioinformatics, physicochemical, characterization

Introduction
Snake venom is a very complex cocktail of powerful proteins and peptides possessing a variety of biological activities. They are the secretory substances which are synthesized and stored in highly specialized venom glands and are used for defense, in an evolutionary adaptation to immobilize the prey. It also helps in digestion of variety of diets of snakes [1]. The use of snake venom in different pathophysiological conditions has been referred in homeopathy and folk medicine for centuries. More than sixty years it was realized that the physiologically active components of snake venoms might have therapeutic potential [2]. Therefore, the venom enzymes are considered as important biological resource and very prominent models for biochemical and structural studies as they offer interesting strategies for vaccines or antidotes design. Several major human drugs or diagnostics have been developed based on snake venom components as well some fundamental biological processes have been disclosed using toxins as probes to study cells and their receptors. Captopril, an antihypertensive drug, is a classic example that was modeled from *Bothrops jararaca* venom [3]. Snake venom may be a crucial nominee for the medicine in the future which could open the doors for new era of medicines and research for treatment of cancer [4-7]. One of the major components of snake venom is the snake venom hemorrhagic metalloproteinases (SVMPs) which found in most Crotalid and Viperid venoms [8]. They belong to the family of zinc endopeptidases varying in size from 20 to 100 kDa, capable of degrading protein of extra cellular matrix and components of hemostatic system [9]. Metalloproteinases of different sizes are often present in the same venom. SVMPs are grouped into several types (PI to P-IV) according to their domain organization that are present in these complexes [10]. They generate

variable tissue and cellular pathology in bite victims. These enzymes induce profuse hemorrhage, blood coagulation, and the inactivation/activation of complement proteins [11]. Nevertheless other members of the SVMPs hold fibrin(-ogen) olytic activity [12], function as prothrombin activators [13] include proapoptotic activity [14], serve as in activators of blood serine proteinase inhibitors [15], and activate factor X (a pro-coagulation protein) [16]. Snake venom encloses panoply of metalloproteinases with miscellaneous biochemical activities. They can be used as potential tools for treating a variety of human disorders and conditions. SVMPs stand for promising candidate drugs for treating diseases concerning abnormal blood clot formation and for reversing brain pathology as observed for some neurodegenerative disorders, hemostatic disorders and local tissue damage since they serve as defibrinogenating agents [17]. In our present study we focused on the investigation of diversification profile of amino acid sequences, secondary structure analysis and conservation pattern of amino acid residues and phylogenetic tree of snake venom metalloproteinase proteins from different snake species. This also helps us to analyze comparative physicochemical and structural properties of snake metalloproteinase and also better understanding of effective conserved motif structure of metalloproteinases.

Materials and Methods

A set of 16 metalloproteinases protein sequences (Table 1) were retrieved in FASTA format [18-20] from National Centre for Biotechnology Information (NCBI), cited at http://www.ncbi.nlm.nih.gov.

Table 1: Metallproteinase sequences retrieved from NCBI database

S. No	Accession no.	Organism's name
01	ABN72536.1	*Bungarus fasciatus*
02	B8K1W0.1	*Daboia russellii russellii*
03	ABN72547.1	*Naja atra*
04	AAQ63966.1	*Protobothrops jerdonii*
05	ABM87941.1	*Ophiophagus hannah*
06	ADI47712.1	*Echis coloratus*
07	JC8020	*Trimeresurus jerdonii*
08	ABE73077.1	*Gloydius halys*
09	P0C6B6.1	*Trimeresurus albolabris*
10	A8QL49.1	*Bungarus multicinctus*
11	ABK63559.1	*Demansia vestigiata*
12	F8RKW1.1	*Drysdalia coronoides*
13	ADI47604.1	*Echis carinatus sochureki*
14	JAB52758.1	*Micrurus fulvius*
15	Q10749.3	*Naja mossambica*
16	AAM27043.1	*Crotalus molossus molossus*

Each Sequence depicts the metalloproteinase from each different snake species. The sequence was used as the input data type to compute various physical and chemical parameters, such as molecular weight, isoelectric point (pI), instability index, aliphatic index and Grand Average of Hydropathy (GRAVY) with the help of Expasy Protparam tool and these properties can be deduced from a protein sequence. The computed Isolelectric point (pI) will be helpful for developing buffer systems for purification by isoelectric focusing method [21]. Estimation of protein stability is possible through instability index where a protein whose instability index is smaller than 40 is predicted stable in nature and a value above 40 predicts that the protein may be unstable in nature [22]. The aliphatic index (AI) of a protein is defined as the relative volume occupied by aliphatic side chains helps to study thermo stable properties of an enzyme. Proteins with a high aliphatic index are more thermally stable. The secondary structural feature was predicted by SOPMA (Self-Optimized Prediction Method with Alignment [23], tool of NPS (Network Protein Sequence Analysis) [24]. This method determines the content of α-helix, β- sheets, turns, random coils and extended strands. SOPMA is a neural network based methods; global sequence prediction may be done by this sequence method [25]. Motif election is very necessary for predicting probable domain of selected metalloproteinases. Motif election and domain analysis was done through MEME. All metalloproteinase sequences from different snake species were subjected to multiple sequence alignment by Clwstal W2. Phylogenetic analyses based on protein sequences were found by using the maximum-likelihood method with MEGA 5.2.2 version.

Results

The metalloproteinases sequences of selected sixteen snake species were retrieved from protein database of NCBI. The physicochemical parameters signify the behavior of metalloproteinases in different condition (Table 2).

Table 2: Physicochemical properties analysis by ProtParam

Accession no.	Molecular weight	Theoretical pI	Grand average of hydropathicity (GRAVY)	Instability index	Aliphatic index
ABN72536.1	68201.7	8.74	-0.548	42.44	66.38
B8K1W0.1	69554.9	5.95	-0.438	37.66	73.20
ABN72547.1	69402.5	5.25	-0.411	39.88	70.98
AAQ63966.1	54613.4	5.27	-0.524	40.50	70.70
ABM87941.1	69048.5	5.67	-0.473	44.17	69.23
ADI47712.1	54992.6	5.43	-0.496	33.91	70.28
JC8020	54613.4	5.27	-0.524	40.50	70.70
ABE73077.1	53141.9	5.31	-0.375	34.99	78.07
P0C6B6.1	54264.0	5.18	-0.477	35.93	73.70

A8QL49.1	68988.1	8.73	-0.487	42.57	73.68
ABK63559.1	68312.2	5.55	-0.472	43.97	68.09
F8RKW1.1	68239.7	5.20	-0.495	44.10	67.24
ADI47604.1	27813.6	5.03	-0.615	44.49	44.98
JAB52758.1	69037.0	8.54	-0.542	41.63	67.70
Q10749.3	68176.1	6.30	-0.542	46.51	64.83
AAM27043.1	46644.8	5.24	-0.387	36.50	88.72

The physicochemical properties illustrate that the molecular weight ranges from ranges from 27813.6 to 69554.9 Da. The highest molecular weight was found in *Daboia russellii russellii (*69554.9 Da) and lowest in *Protobothrops jerdonii* (27813.6 Da). The computed Isoelectric point (pI) of the studied metalloproteinases revealed their acidic behavior (pI<7), whereas pI for *Bungarus fasciatus* (ABN72536.1), *Bungarus multicinctus* (A8QL49.1), *Micrurus fulvius (*JAB52758.1) indicated their basic behaviour. On the basis of instability index Expasy's ProtParam classified the ten metalloproteinases were unstable (Instability index >40) and remaining six *Daboia russellii russellii* (B8K1W0.1), *Naja atra* (ABN72547.1), *Echis coloratus* (ADI47712.1), *Gloydius halys* (ABE73077.1), *Trimeresurus albolabris* (P0C6B6.1) and *Crotalus molossus molossus* (AAM27043.1) respectively were stable (Instability index <40). In addition, Aliphatic Index (AI) of studied metalloproteinases found to span within a range of 64.83 to 88.72. Stability of snake metalloproteinases in a small range suggests their unstable nature over wide temperature range, although particular metalloproteinase of *Crotalus molossus molossus* is observed as the most thermo stable metalloproteinase. GRAVY value of the considered metalloproteinases demonstrated that all proteins show lower GRAVY value which indicates the better interaction of that protein with water. The very low GRAVY index of proteins from *Crotalus molossus molossus* (AAM27043.1) and *Gloydius halys* (ABE73077.1) infers that these proteins could result in a better interaction with water.

Secondary structural analysis indicates that random coil predominate the other structures where β-turn being the least conformational structure (Table 3). Among all studied metalloproteinase, it was found that β-turns showing very less percentage of conformation (below 13%). Extended strands were ranging from 19- 28% in the majority of the metalloproteinase. The secondary structure shows whether a given amino acid lies in a helix, strand or coil [26-27].

Table 3: Secondary structure analysis by SOPMA

Accession no.	Alpha helix (Hh %)	Extended strand (Ee %)	Beta turn (Tt %)	Random coil (Cc %)
ABN72536.1	22.31	25.62	12.89	39.17
B8K1W0.1	26.18	25.85	9.11	38.86
ABN72547.1	*23.67*	24.96	10.79	40.58
AAQ63966.1	25.62	23.97	9.09	41.32
ABM87941.1	24.88	24.71	9.49	40.92
ADI47712.1	*21.26*	22.27	10.53	45.95
JC8020	25.62	23.97	9.09	41.32
ABE73077.1	26.42	22.64	9.64	41.30
P0C6B6.1	21.49	24.59	9.09	44.83
A8QL49.1	21.82	25.57	10.59	42.02
ABK63559.1	21.60	25.70	9.49	43.21
F8RKW1.1	16.97	26.59	8.65	47.80
ADI47604.1	14.57	19.43	10.93	55.06
JAB52758.1	20.72	28.22	9014	41.92
Q10749.3	22.82	24.14	11.49	41.54
AAM27043.1	24.40	27.29	10.14	38.16

Phylogenetic analyses based on protein sequences were done by the maximum likelihood method with the MEGA 5.2.2. The resulting tree is represented in (Fig. 1).

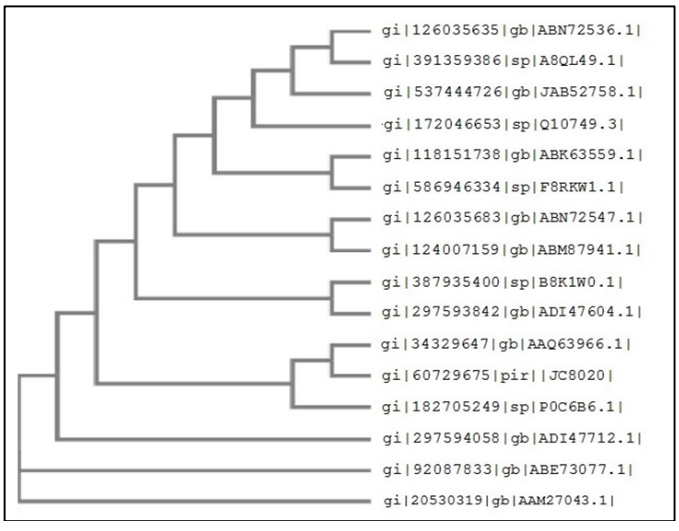

Fig 1: phylogenetic tree of neurotrxin

Phylogenetic tree was built using Neighbor-Joining method. A number of clusters with close relationships were revealed including *Bungarus fasciatus* (ABN72536.1) and *Bungarus multicinctus* (A8QL49.1), *Demansia vestigiata* (ABK63559.1) *and Drysdalia coronoides* (F8RKW1.1), *Naja atra (*ABN72547.1) *and Ophiophagus hannah (*ABM87941.1). Furthermore, *Daboia russellii russellii (*B8K1W0.1) *and Echis carinatus* (ADI47604.1), *Protobothrops jerdonii (*AAQ63966.1) *and Trimeresurus jerdonii (*JC8020) has close relation *Trimeresurus albolabris* (P0C6B6.1). Proteins in close evolutionary relationship may be considered together for their association in similar biological processes.

Motif analysis of the metalloproteinase sequences was performed by using MEME. The output of this modal of MEME shows color graphical alignment with common regular expression of motifs. The block depicts start and end point of the amino acid sequences with motif length. This is recognized fact E-value describes the statistical significance of the motif. The E-value is an evaluation of the probable number of motifs with the given log probability ratio (or higher) along with the same width and site count present, that one would find in a similarly sized set of random sequences. On the other hand, motif width identifies that each motif describes a pattern of a fixed with as no gaps are allowed in MEME motifs. In MEME package, sites characterize the conserved regions present in the particular motifs. Site numbers are the important contributing factor to the construction of the motifs. MEME selects the width and number of occurrences of each motif automatically to minimize the 'E-value' of the motif which increases the probability of finding an equally well-conserved pattern in random sequences. Motif overview has revealed in figure describing 1.5e-440 E-value of motif one, 7.0e-572 E-value of motif two and 82.9e-516 Evalue of motif three. E-value, width, sites and sequence logo are given in Table 4.

Table 4: MEME result of 3 motifs of metalloproteinase sequences.

Motif number	Width	E-Value	Sites	Sequence logo
1	50	1.5e-440	13	
2	50	7.0e-572	15	
3	50	2.9e-516	15	

These three motifs (Fig. 2) found by MEME are subjected to Pfam (protein family database) to identify the domain of protein family related to our MEME motifs. Multiple sequence alignment using ClustalW2 shows significant alignment pattern (Fig. 3).

Motif 1 sites sorted by position p-value

Sequence name	Start	P-value	Site
gi|182705249|sp|P0C6B6.1|	1	2.19e-60	. MIQVLIVTICLAVFPYQGSSIILESGNVNDYEVVYPEKVTALPKGAVQQK YEDAMQYEFK
gi|391359386|sp|A8QL49.1|	1	5.48e-60	. MIQALIVTICLAVFPYQGSSIILESGNVNDYEVVYPQKVPLLPKGGVQNP QPKTKYEDTV
gi|118151738|gb|ABK63559.1|	1	9.12e-60	. MIQALIVTICFMVFPYQGSSIILESGNVNDYEVVYPQKVPALPKGGVQNP QPETKYEDTM
gi|172046653|sp|Q10749.3|	1	7.54e-59	. MIQALIVAICLAVFPYQGSSIILESGNVNDYEVVYPQKVPALSKGGVQNP QPETKYEDTM
gi|60729675|pir||JC8020|	1	9.58e-59	. MIQVLIVTICLAVFPYQGSSIILESGNIDDYEVVYPRKVTALPKGAVQQK YEDTMQYEFK
gi|34329647|gb|AAQ63966.1|	1	9.58e-59	. MIQVLIVTICLAVFPYQGSSIILESGNIDDYEVVYPRKVTALPKGAVQQK YEDTMQYEFK
gi|586946334|sp|F8RKW1.1|	1	3.50e-58	. MIEVLIVTICFTVFPYQGSSIILESGNVNDYEVVYPQKVPALPKGGVQNP QPETKYEDTM
gi|387935400|sp|B8K1W0.1|	1	2.01e-55	. MMQVLIVTICLAVFPYHGSSIILESGNVNDYEVVYPQKVTAMPKEAVKQP EQKYEDAMQY
gi|537444726|gb|JAB52758.1|	1	5.61e-55	. MIQALIVTICLAVFPYKGSSIVLESGNVNDYEVVYPQKVPLLPKRGVQNP QPETKYEDTM
gi|20530319|gb|AAM27043.1|	1	3.43e-54	. MIEIVLVTICLAVFPCQGSSIILESGNVNDYEVVYPRKVTALPKGAVQPK YEDAMQYELK
gi|124007159|gb|ABM87941.1|	1	4.34e-53	. MIQVLIVTICLVVFPYQGSSIILESGKVNDYEVVYPQKIPVLPKSKIQRR EQKMYEDTMK
gi|126035683|gb|ABN72547.1|	1	8.62e-50	. MIQPLIVAICLVVFPYQGSSTILESGKVRDYEVVYPQKIPSLPKGRLQRR EEKTKYENTM
gi|297594058|gb|ADI47712.1|	1	8.16e-49	. MTQVLIVTICLAVFPYQGSSKTLKSGNVNDYEVVNPQKITGLPVGAFKQP EKKHEDAVQY

Motif 2 sites sorted by position p-value

Sequence name	Start	P-value		Site	
gi|586946334|sp|F8RKW1.1|	66	2.50e-63	YEDTMQYEFH	VNGEPVVLHLERNKGLFSEDYTETHYAPDGREITTSPPVQDHCYYHGYIQ	NEADSSAAIS
gi|118151738|gb|ABK63559.1|	66	1.16e-61	YEDTMQYKFH	VNGEPVVLHLERNKCLFSEDYTETHYADDGREITTSPPVQDHCYYHCSIQ	NEADSSAVIS
gi|124007159|gb|ABM07941.1|	65	2.72e-61	YEDTMKYEFK	VNGEPVVLHLERNKELFSKDYTETHYSPDGREITTSPPVEDHCYYHGYIQ	SDIDSTAILN
gi|172046653|sp|Q10749.3|	66	4.16e-61	YEDTMQYEFH	VNGEPVVLHLERNKGLFSEDYTETHYAPDGREITTSSPVQDHCYYHGYIQ	NEADSSAVIS
gi|391359386|sp|A8QL49.1|	66	1.28e-60	YEDTVQYEFE	VNGEPVVLHLERNKGLFSEDYTEAHYAPDGREITTRPPVQDHCYYHGYIQ	NEADSSAAIS
gi|60729675|pir||JC8020|	61	6.64e-60	YEDTMQYEFK	VNEEPVVLHLEKNKGLFSKDYSETHYSPDGREITTYPPVEDHCYYHGRIQ	NDADSTASIS
gi|34329647|gb|AAQ63966.1|	61	6.64e-60	YEDTMQYEFK	VNEEPVVLHLEKNKGLFSKDYSETHYSPDGREITTYPPVEDHCYYHGRIQ	NDADSTASIS
gi|92087833|gb|ABE73077.1|	60	2.18e-59	YEDAMQYELK	VNGEPVVLHLEKNKGLFSKDYIETHYSPDGRKITTNPPVEDHCYYHGRIQ	NDADSTASIS
gi|297594058|gb|ADI47712.1|	64	2.75e-59	HEDAVQYEFE	VNGEPVVLHLEKNKGLFSEDYSETHYSPDGREITTNPPVEDHCYYHGRIK	NDADSTASIS
gi|126035683|gb|ABN72547.1|	66	8.47e-59	YENTMKYEFK	VNGEPVVLNLEKNKRLFSKDYTETHYSPDGREITTSPPVQDHCYYHGHIQ	NDADSTAVIR
gi|537444726|gb|JAB52758.1|	66	4.34e-58	YEDTMQYEFQ	MKGEPVVLHLERNKGLFSEDYTETHYAPDGRIITTSPPVQDHCYYHGYIQ	NDADSSAVIS
gi|387935400|sp|B8K1W0.1|	64	9.47e-57	YEDAMQYKFE	VNGEPVLLHLEKNKDLFSEDYSETHYSPDGREITTKPLVQDHCYYHGHIQ	NDAHSSASIS
gi|126035635|gb|ABN72536.1|	57	1.75e-56	YEDTMQYEFH	VNGEPVVPHLEKNKGLFSEDYTETHYASDGREITTSPSVQDHCYYYGYIQ	NEADSSAAIS
gi|20530319|gb|AAM27043.1|	61	1.60e-54	YEDAMQYELK	VNGEPVVLHLEKNKELFSKDYSETHYSPDGRKITTNPLVEDHCYYRGRIE	NDADSTASIS
gi|182705249|sp|P0C6B6.1|	61	2.13e-54	YEDAMQYEFK	VNGEPVVLYLEKNKELFSENYSETHYSLDGREITTYPSVEDHCYYHGRIQ	NDADSTASIS

Motif 3 sites sorted by position p-value

Sequence name	Start	P-value		Site	
gi|60729675|pir||JC8020|	113	6.97e-61	CYYHGRIQND	ADSTASISACNGLKGHFKLQGETYFIEPLKLPDSEAHAVFKYENVEKEDE	APKMCGVTET
gi|34329647|gb|AAQ63966.1|	113	6.97e-61	CYYHCRIQND	ADSTASISACNCLKCHFKLQCETYFIEPLKLPDSEAHAVFKYENVEKEDE	APKMCCVTET
gi|118151738|gb|ABK63559.1|	110	7.67e-59	CYYHGSIQNE	ADSSAVISACDGLKGHFKHRGETYFIEPLKLSDSESHAIYKDENVEKEDE	IPQICGVTHT
gi|126035635|gb|ABN72536.1|	109	1.25e-58	CYYYGYIQNE	ADSSAAISACDGLKGHFKHQGETYFIEPLKLSNSEAHAIYKDENVEEEDE	TPKICGLTQT
gi|391359386|sp|A8QL49.1|	118	1.21e-57	CYYHGYIQNE	ADSSAAISACDGLKGHFKHQGETYFIEPLKLSDSEAHAIYKDENVEEENE	TPKICGLTET
gi|92087833|gb|ABE73077.1|	112	1.21e-57	CYYHGRIQND	ADSTASISACNGLKGHFKLQGETYLIEPLKLSNSEAHAVYKYEDVEKEDE	APKMCGVTQN
gi|537444726|gb|JAB52758.1|	118	1.53e-57	CYYHGYIQND	ADSSAVISACDGLKGHFKHQGETYFIEPLKLSDSEAHAIYKDENVEKEDV	TLKICGVTQT
gi|387935400|sp|B8K1W0.1|	116	5.91e-57	CYYHCHIQND	AHSSASISACNCLKCHFKLRCEMYLIEPLKLSDSEAHAVYKYENVEKEDE	ALKMCCVTQT
gi|586946334|sp|F8RKW1.1|	118	2.72e-56	CYYHGYIQNE	ADSSAAISACDGLKGHFKHRGETYFIEPLKLSNSESHAIYKDEHVEKEDE	IPKICGVTQT
gi|20530319|gb|AAM27043.1|	113	3.74e-56	CYYRGRIEND	ADSTASISTCNGLKGHFKLQGEIYLIEPLKLSDSEAHAVFKLENVEKEDE	GPKMCGVTQN
gi|182705249|sp|P0C6B6.1|	113	7.03e-56	CYYHGRIQND	ADSTASISACNGLKGHFKLRGKTYLIEPLKLPNSEAHAVFKYENVEKEDE	APQMCGVTET
gi|172046653|sp|Q10749.3|	118	2.69e-55	CYYHGYIQNE	ADSSAVISACDGLKGHFKHQGETYFIEPLELSDSEAHAIYKDENVEEEEE	IPKICGVTQT
gi|126035683|gb|ABN72547.1|	118	6.53e-49	CYYHGHIQND	ADSTAVIRACDGLNGYFKSNGEMYIIEPLKLSDSEAHAVFKYESLEKEDE	TPKTCGAIHN
gi|124007159|gb|ABM87941.1|	117	1.07e-48	CYYHGYIQSD	IDSTAILNACNGLKGYFRHHGEAYHIEPLKFSDSEAHAVYKYENIEKEDE	TPKICGVKHS
gi|297594058|gb|ADI47712.1|	116	8.65e-47	CYYHGRIKND	ADSTASISTCNGLKGFFILRGETYLIEPLKVPDSEAHAVYKYEDGKKDE	APKMCGVTQT

Fig 2: Motif Elkection by me me modf1, modf2, motif3

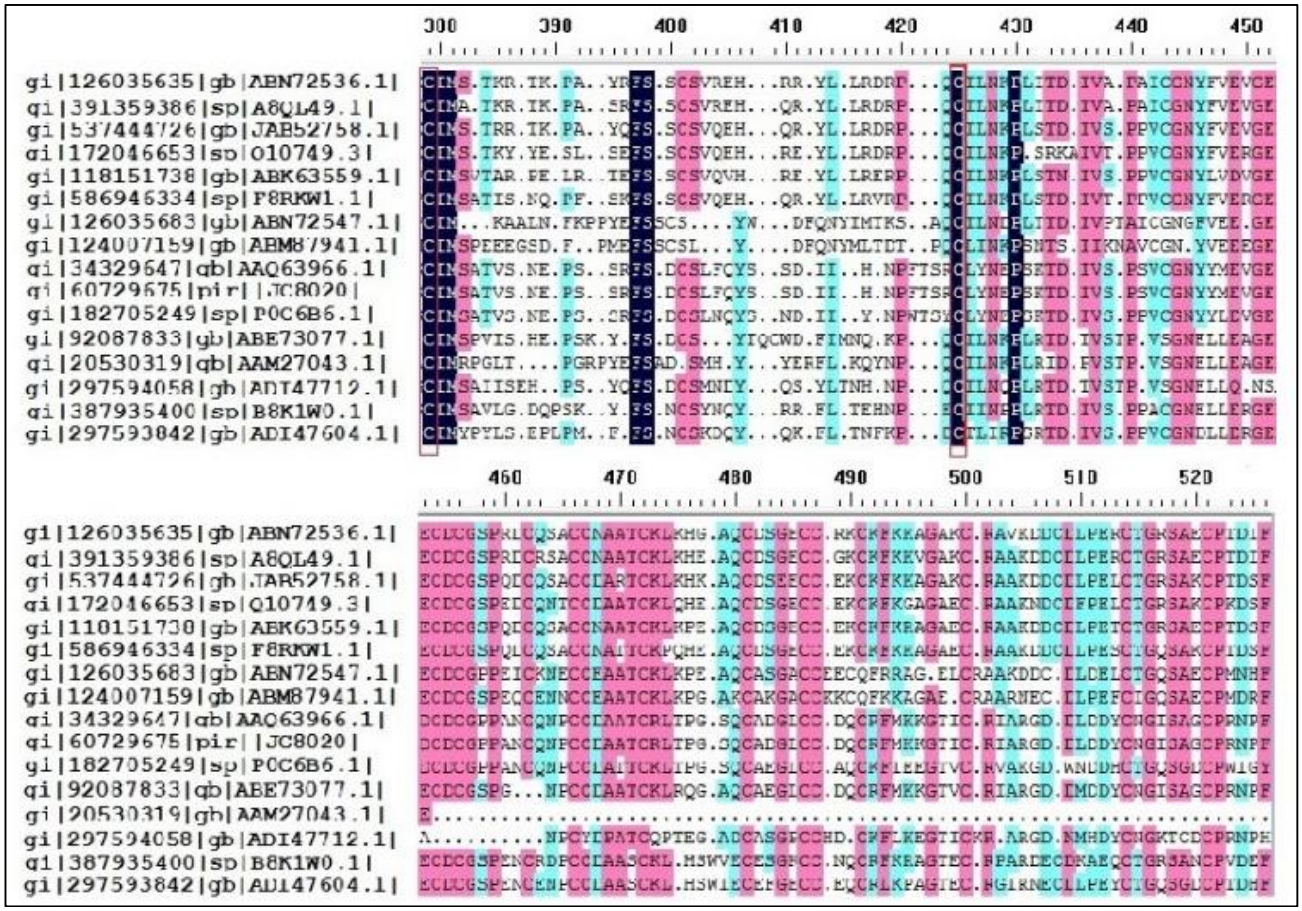

Fig 3: Moultiple Sequence Alignment by Clustal W2

Discussion

The physicochemical properties of metalloproteinases stand for several features which reveal molecular weight, theoretical pI, instability index, aliphatic index and grand average of hydropathicity. In this investigation, most studied metalloproteinases are unstable, acidic and can exhibit better interaction with water. A predominance of random coil is seen in the secondary structure of metalloproteinases. This is proficient to predict three dimensional structures of proteins and can assist in approximation of some aspects of protein function and their classification into families [28-29]. Phylogeny stands for the evolutionary spectrum of studied metalloproteinases. Motif is very significant for representing the domain of particular protein family. There are some conserved regions in a protein sequence which are known as motifs. Each protein has specific function in the living organism and proteins can be characterized by more than one motif and it can be classified using certain specific motifs. Motif helps to observe the functional domain of proteins and also motif represents the conserved pattern in protein sequences through which we can design degenerate primer of that protein sequence. Motif one embodies the similarity with Domain of Unknown Function (DUF) protein family and more than 30% of all protein domains are currently annotated as "Domains of Unknown Function" (DUFs). Evolutionary conservation recommends that many of these DUFs are very significant [30]. All motifs (one, two and three) represent the domain of reprolysin protein family. By using ClustalW2 tool, multiple sequence alignment was done which displays a useful alignment pattern where all motifs were sought out. Comparative analysis by using multiple sequence alignment, conserved cysteine is present in all sequences and cysteine also finds in maximum sequences. Cysteine amino acid illustrates the disulphide bonds and also reveals the subsistence of functional domain.

Conclusion

In this research, it was tried to disclose the hidden information about 16 metalloproteinases sequences from different snake species by analyzing their structural features e.g. physico-chemical properties, secondary structural features and phylogenetic classification. Intensive characterization and comparative analysis of the neurotoxins with the help of numerous bio-computational tools were used to ease up the process of finding and yielded novel insights and views which can be employed in pharmaceutical industry. Secondary structure studies showed that all the studied proteins contain high percentage of random coils and the least conformation was of β-turns. In motifs conserved sequences aid us to culminate an important insight of functional domain. Conserved sequences also may be exploited for designing specific degenerate primers for identification and isolation of type and class of neurotoxins as numerous neurotoxins are being isolated to fulfill the need of efficient application in various system. This study has shown the preliminary outlook of metalloproteinases. More study is needed which will facilitate us to comprehend the structure and mechanism of metalloproteinases. The findings throughout this study may be used by researchers working on snake metalloproteinases in context of any experimental system.

Acknowledgements

Authors are thankful to Dr. Md. Abu Reza, (Professor of Genetic Engineering and Biotechnology Department, University of Rajshahi, Rajshahi-6205, Bangladesh) for his valuable support and discussions related to this research.

References

1. Kang TS, Georgieva D, Genov N, Murakami MT, Sinha M, Kumar RP, Kaur P, Kumar S, Dey S, Sharma S, Vrielink A, Betzel C, Takeda S, Arni RK, Singh TP, and Kini R. Enzymatic toxins from snake venom: Structural characterization and mechanism of catalysis. FEBS J, 2011; 278(23):4544-4576.
2. Pal SK, Gomes A, Dasgupta SC, Gomes A. Snake venom as therapeutic agents: from toxin to drug development. Indian J Exp Biol. 2002; 40:1353-1358.
3. Ferreira SH, Bartelt DC, Greene LJ. Isolation of bradykinin-potentiating peptides from *Bothrops jararaca* venom. Biochem, 1970; 9(13):2583-2593.
4. Park MH, Choi MS, Kwak DH, Oh K, Yoon DY, Han SB, Song HS, Song MJ, and Hong JT. Anti-cancer effect of bee venom in prostate cancer cells through activation of caspase pathway via inactivation of NF- κB. Prostate, 2011; 71(8):801-812.
5. Tang N, Xie Q, Wang X, Li X, Chen Y, Lin X, and Lin J. Inhibition of invasion and metastasis of MHCC97H cells by expression of snake venom cystatin through reduction of proteinases activity and epithelial-mesenchymal transition. Arch Pharm Res. 2011; 34(5):781-789.
6. Heinen TE, and da Veiga AB. Arthropod venoms and cancer. Toxicon. 2011; 57(4):497-511.
7. Islam M, Roly ZY, Ruhullah M, Islam Z, Sharia A, Tanvir RZ and Foysal K. Identification of accelerated evolution in the metalloproteinase domain of snake venom metalloproteinase sequences (SVMPs) through comparative analysis. Afr. J Biotechnol. 2016; 15(8):252-263.
8. Markland FSJr, Swenson S. Snake venom metalloproteinases. Toxicon. 2013; 62(1):3-18.
9. Panfoli I, Calzia D, Ravera S, Morelli A. Inhibition of haemorrhagic snake venom components: old and new approaches. Toxins, 2010; 2(4):417-427.
10. Hite LA, Jia LG, Bjarnason JB, Fox JW. cDNA sequences for four snake venom metalloproteinases: structure, classification, and their relationship to mammalian reproductive proteins. Arch Biochem Biophys. 1994; 308(1):182-191.
11. Rael ED, Rivas JZ, Chen T, Maddux N, Huizar E, Lied CS. Differences in fibrinolysis and complement inactivation by venom from different northern blacktailed rattlesnakes (Crotalus molossus molossus). Toxicon, 1997; 35(4):505-513.
12. Markland FS. Fibrolase, an active thrombolytic enzyme in arterial andvenous thrombosis model systems. Adv Exp Med Biol. 1996; 391(2):427-438.
13. Kini RM. The intriguing world of prothrombin activators from snake venom. Toxicon. 2005; 45(8):1133-1145.
14. Araki S, Masuda S, Maeda H, Ying MJ, Hayashi H. Involvement of specific integrins in apoptosis induced by vascular apoptosis-inducing protein. Toxicon. 2002; 40(5):535-542.
15. Kress LF, Kurecki T. Studies on the complex between human alpha 2-macroglobulin and Crotalus adamanteus proteinase II. Release of active proteinase from the complex. Biochim Biophys Acta. 1980; 613(2):469-475.
16. Siigur E, Aaspõllu A, Trummal K, Tõnismägi K, Tammiste I, Kalkkinen N, *et al*. Factor X activator from Vipera lebetina venom is synthesized from different genes. Biochim Biophys Acta. 2004; 1702(1):41-51.
17. Costa Jde O, Fonseca KC, Garrote-Filho MS, Cunha CC, de Freitas MV, Silva HS, *et al*. Structural and functional comparison of proteolytic enzymes from plant latex and snake venoms. Biochimie. 2010; 92(12):1760-65.
18. Roly ZY, Islam MM, and Reza MA. A comparative *in silico* characterization of functional and physico-chemical properties of 3FTx (three finger toxin) proteins from four venomous snakes. Bioinformation. 2014; 10(5):281-287.
19. Islam MM, Haque ME, Dutta AK, Islam MA, Khalekuzzaman M, Sikdar B. A computational study on metalloproteinase in the snake venomous of different rattle snake species in comparison with the asian snake species. Int. J. Pharm. Bio. Sci. 2013; 4(4):741-748.
20. Islam Z, Islam MM, Saha S, Jahangir CA, Basak B, Islam MN, *et al*. Identification and computational analysis of chicken alpha-1 collagen sequences. Int. J Sci. Eng. Res. 2015; 6(1):217-221.
21. Sivakumar K, Balaji S, Gangaradhakrishnan. *In silico* characterization of antifreeze proteins using computational tools and servers. J Chem Sci. 2007; 119(5):571-579.
22. Guruprasad K, Reddy BVB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng. 1990; 4(2):155-161.
23. http://npsapbil.ibcp.fr/cgicgibin/npsa_automat.pl?page=/NPSA/npsa_sopma.html
24. Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci. 1995; 11(5):681-684.
25. Prashant VT, Uddhav SC, Madura SM, Vishal PD, Renuka RK. Secondary structure prediction and phylogenetic analysis of salt tolerant proteins. Global J Molecular Sci. 2014; 5(1):30-36.
26. Jyotsna C, Ashish P, Shailendra G, Verma MK. Homology modeling and binding site identification of 1 deoxy d- xylulose 5 phosphate reductoisomerase of Plasmodium falciparum: New drug target for Plasmodium falciparum. Int J of Engin Sci and Tech, 2010; 2(8):3468-3472.
27. Ojeiru FE, Kazuya T, Yuki H, Mohammed SM, Shunsuke M. Circular dichroism studies on C-terminal zinc finger domain of transcription factor GATA-2. Yonago Acta medica. 2010; 53(1):25-28.
28. Rost B. Review: protein secondary structure prediction continues to rise. J Struct Biol. 2001; 134(2):204-218.
29. Roly ZY, Ruhullah M, Tanvir RZ, Foyasal MK, Bapari MR, Roy D, *et al*. Functional Characterization of Elapid Venom Toxin Proteins and Related Non-Toxin Proteins of Human through Comparative Analysis. Trends in Bioinformatics. 2015; 8(3):86-92.
30. Goodacre NF, Gerloff DL, Uetz P. Protein domains of unknown function are essential in bacteria. MBio. 5(1), e00744-13.