TANG HUMANITAS MEDICINE

Original Article

Characterization of *Bacillus anthracis* proteases through protein-protein interaction: an *in silico* study of anthrax pathogenicity

Amrita Banerjee[1], Shilpee Pal[1], Tanmay Paul[1], Keshab Chandra Mondal[1], Bikash Ranjan Pati[1], Arnab Sen[2], Pradeep Kumar Das Mohapatra[1,*]

[1]*Bioinformatics Infrastructure Facility Centre, Department of Microbiology, Vidyasagar University, Midnapore-721102, West Bengal, India;* [2]*Department of Botany, Molecular Genetics Laboratory, School of Life Sciences,University of North Bengal, Siliguri734013, India*

ABSTRACT

Anthrax is the deadly disease for human being caused by *Bacillus anthracis.* Instantaneous research work on the mode of infection of the organism revealed that different proteases are involved in different steps of pathogenesis. Present study reports the *in silico* characterization and the detection of pathogenic proteases involved in anthrax infection through protein-protein interaction. A total of 13 acid, 9 neutral, and 1 alkaline protease of *Bacillus anthracis* were selected for analysing the physicochemical parameter, the protein superfamily and family search, multiple sequence alignment, phylogenetic tree construction, protein-protein interactions and motif finding. Among the 13 acid proteases, 10 were found as extracellular enzymes that interact with immune inhibitor A (InhA) and help the organism to cross the blood brain barrier during the process of infection. Multiple sequence alignment of above acid proteases revealed the position 368, 489, and 498-contained 100% conserved amino acids which could be used to deactivate the protease. Among the groups analyzed, only acid protease were found to interact with InhA, which indicated that metalloproteases of acid protease group have the capability to develop pathogenesis during *B. anthracis* infection. Deactivation of conserved amino acid position of germination protease can stop the sporulation and germination of *B anthracis* cell. The detailed interaction study of neutral and alkaline proteases could also be helpful to design the interaction network for the better understanding of anthrax disease.

Keywords Anthrax, *Bacillus anthracis*, protease, superfamily and family, phylogenetic tree**,** motif, protein-protein interaction

INTRODUCTION

Protease, renowned as proteolytic enzymes or proteinases, refers to a group of enzymes that hydrolyzes (breakdown) proteins into small peptides and amino acids. Proteolytic enzymes are essential in various therapeutic purposes such as oncology, inflammatory conditions, blood rheology control, and immune regulation. Undigested proteins, cellular debris and blood toxins can also be digested by proteases. Current classification of enzymes unveils six broad groups of proteases, i.e. serine proteases, threonine proteases, cysteine proteases, aspartate proteases, glutamic proteases and metalloproteases. Alternatively, on the basis of the isoelectric points (pI) of catalytic proteins they can also be classified into acid, alkaline and neutral proteases. Acid and neutral proteases are involved in type I hypersensitivity by activating complement systems and kinins (Mitchell et al., 2007) and the function of alkaline or basic proteases of *Bacillus anthracis* is still unknown.

Bacillus anthracis, pathogens of anthrax, is a gram positive, endospore-forming, rod-shaped bacterium with 1 - 1.2 µm in width and the etiologic agent of anthrax disease. According to previous study (Russell et al., 2007) the mode of invasion of *B. anthracis* can occur in three forms: cutaneous (skin), gastrointestinal (digestive), and inhalation (lungs). The cutaneous anthrax rarely fatal if treated, gastrointestinal anthrax shows 25 - 60% death and the inhalation anthrax is more deadly than others. *B. anthracis* possesses acid, neutral as well as alkaline protease. Some strains such as *B. anthracis* str. CDC 684 contains both acid and alkaline protease gene in their whole genome. The main fatal complication caused by these bacteria is hemorrhagic meningitis but the pathogenesis is still unknown. Mukherjee et al. (2011) showed that *Bacillus anthracis* increases permeability of human brain microvasculature endothelial cells (HBMECs) which constitute the blood-brain barrier (BBB) by secreting metalloprotease InhA on the monolayer integrity of HBMECs. According to Chertow (2011) and Tonry et al. (2012), immune inhibitor a metalloprotease (InhA) of *Bacillus anthracis* helps in adhesion and invasion of human brain endothelial cells by modifying cell surface properties through direct proteolysis of adhesin protein. Anthracis protease cleaves the anthrax lethal protein factor to be internalised by the host cell endocytosis (John, 2010). *B. anthracis* was also used as effective vector for production of recombinant proteins after deletion of six proteases (Pomerantsev et al., 2011). Experimental upshot makes it

*Correspondence: Pradeep Kumar Das Mohapatra
E-mail: pkdmvu@gmail.com
Received November 22, 2013; **Accepted** February 17, 2013; **Published** February 28, 2014
doi: http://dx.doi.org/10.5667/tang.2013.0031

Table 1. List of acid neutral and alkaline proteases with their type and accession numbers which were taken for analysis

No.	Strain name	Type of acid protease	Accession No.	No.	Strain name	Type of acid protease	Accession No.
1	str. A0248	neutral protease B	YP_002869281.1	1	str. BF1	intracellular serine protease	EJY93592.1
2	str. CDC684	neutral protease Npr599	YP_002816538.1	2	str. BF1	germination protease	EJY92062.1
3	str. Sterne	germination protease	YP_030468.1	3	str. BF1	Protease	EJY93586.1
4	str. Sterne	neutral protease B	YP_031148.1	4	str. BF1	serine protease	EJY89664.1
5	str. 'Ames Ancestor'	membrane-bound protease	YP_018763.1	5	str. BF1	serine protease	EJY93228.1
6	str. BF1	Neutral protease B , Bacillolysin	EJY90308.1	6	str. BF1	Serine protease, subtilase family protein	EJY89495.1
7	str. BF1	Neutral protease	EJY93035.1	7	str. UR-1	metalloprotease	EJT19501.1
8	str. BF1	Extracellular neutral metalloprotease, bacillolysin	EJY94285.1	8	str. H9401	Serine protease, subtilase family	AFH85759.1
9	str. H9401	Neutral protease B, Bacillolysin	AFH86425.1	9	str. Australia 94	neutral protease	ZP_05212446.1
10	str. H9401	Extracellular neutral metalloprotease, bacillolysin	AFH83987.1	**No.**	**Strain name**	**Type of acid protease**	**Accession No.**
11	str. H9401	Membrane-bound protease	AFH83399.1				
12	str. A1055	neutral protease	ZP_05184332.1	1	str. CDC 684	alkaline serine protease, subtilase family	YP_002814818.1
13	str. A1055	neutral protease Npr599	ZP_05184332.1				

understand that protease enzyme of *Bacillus anthracis* carries an important active site responsible for its lethal effect. Protein sequences analysis of *B. anthracis* protease, may disclose the underling secrets about the functions and evolutionary relatedness.

Previous *in silico* study on different enzymes like tannase from bacterial and fungal origin (Banerjee et al., 2012), alkaline proteases from different species of *Aspergillus* (Morya et al., 2012), xylanase from *Thermomyces lanuginosus* (Shrivastava et al., 2007), pectate lyase from different sources (Dubey et al., 2010), have been reported but *in silico* study on *B. anthracis* protease protein sequences is still unrevealed.

Present work has been designed to understand the natures of different types of *B. anthracis* proteases through *in silico* comparative study. Analysis and characterization of 13 acid proteases, 9 neutral proteases and one alkaline protease were performed. The protein sequences were employed to analyze various physiochemical parameter analyses, super family

Table 2. Physiochemical parameter analysis

No.	Acc No.	Type of protease	Amino acid No.	Molecular weight	Theoretical	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Instability index	Aliphatic index	Grand average of hydropathicity (GRAVY)
Acid										
1	YP_002869281.1	NP(B)	547	60426.4	5.54	77	62	26.41	71.35	-0.527
2	YP_002816538.1	NP(npr599)	566	61004.5	5.82	65	58	30.56	71.20	-0.487
3	YP_030468.1	GP	368	40605.2	4.94	58	39	38.80	96.36	-0.235
4	YP_031148.1	NP(B)	552	60909.0	5.54	77	62	25.90	72.46	-0.505
5	YP_018763.1	MBP	742	86273.7	8.65	70	76	29.35	84.35	-0.208
6	EJY90308.1	NP(B)	549	60639.7	5.61	77	63	26.14	71.44	-0.528
7	EJY93035.1	NP	565	62634.3	5.25	71	56	30.30	75.06	-0.540
8	EJY94285.1	ENMP	567	62253.8	5.29	71	57	25.90	69.14	-0.603
9	AFH86425.1	NP(B)	554	61122.3	5.61	77	63	25.64	72.55	-0.507
10	AFH83987.1	ENMP	581	63931.9	5.50	72	61	25.59	69.98	-0.599
11	AFH83399.1	MBP	635	73571.4	7.24	66	66	28.82	77.07	-0.365
12	ZP_05184332.1	NP	345	38566.2	6.29	50	48	25.48	75.97	-0.690
13	ZP_05183912.1	NP(npr599)	509	55622.6	5.53	64	55	31.95	69.57	-0.551
Neutral										
1	EJY93592.1	ISP	316	34056.5	4.92	43	28	35.51	97.47	-0.191
2	EJY92062.1	GP	368	40605.2	4.94	58	39	38.80	96.36	-0.235
3	EJY93586.1	P	386	43880.3	8.70	45	50	24.13	69.53	-0.659
4	EJY89664.1	SP	391	41818.7	7.75	44	45	27.39	98.98	-0.106
5	EJY93228.1	SP	413	43898.7	5.73	49	46	26.05	89.18	-0.259
6	EJY89495.1	SP	613	67754.7	4.93	80	55	38.57	83.62	-0.465
7	EJT19501.1	MBP	422	48910.5	9.33	33	45	32.73	104.64	0.227
8	AFH85759.1	SP	929	100172.0	5.83	99	87	29.09	86.07	-0.332
9	ZP_05212446.1	NP	545	59343.3	4.98	78	58	20.29	76.83	-0.422
Alkaline										
1	YP_002814818.1	ASP	397	42380.9	8.29	30	32	20.10	75.92	-0.326

NP (B/npr599): Neutral protease B/npr599, GP: Germination protease, MBP: Membrane-bound protease, ENMP: Extracellular neutral metalloprotea se, ISP: Intracellular serine protease, P: protease, SP: serine protease, ASP: Alkaline serine protease.

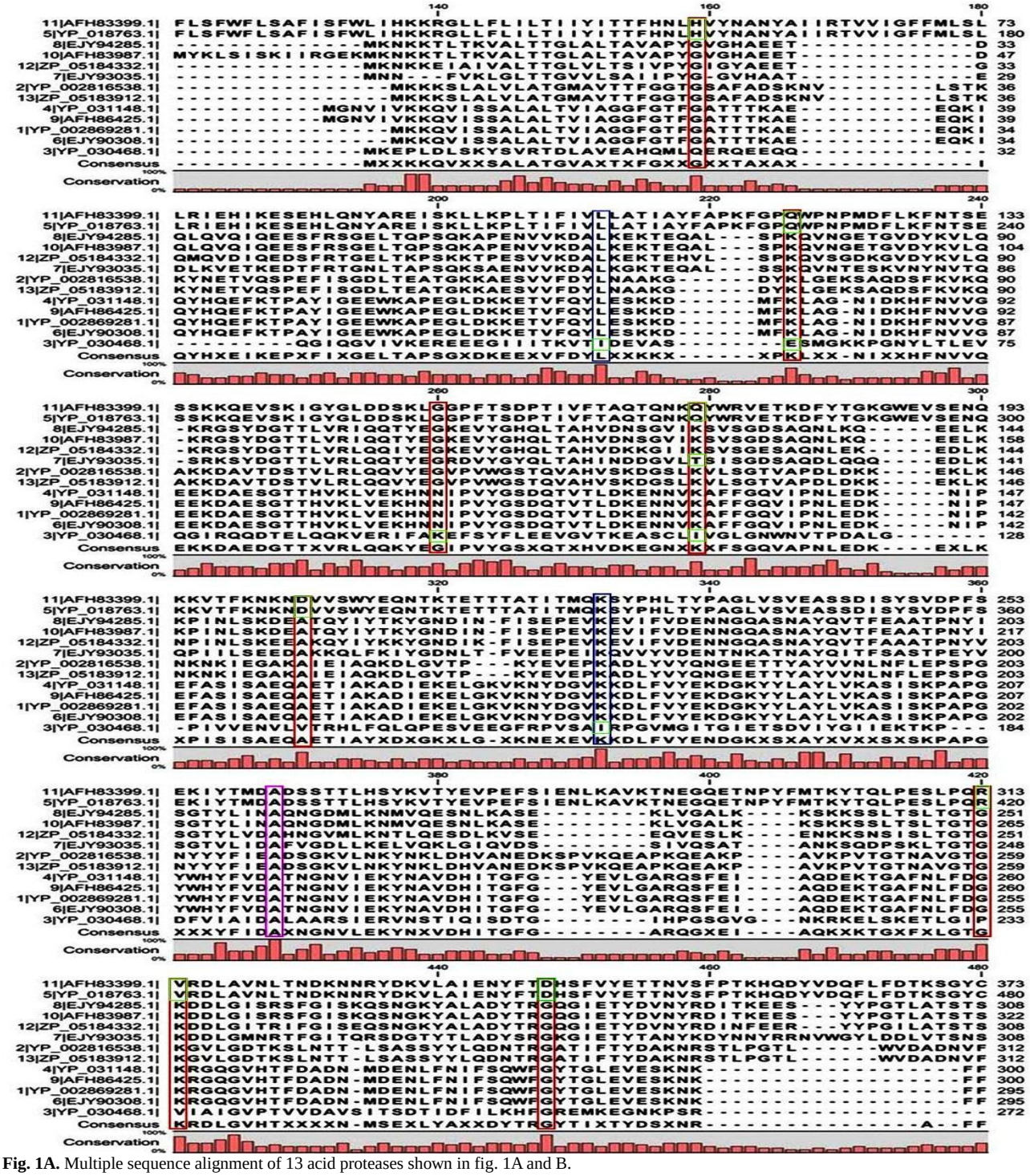

Fig. 1A. Multiple sequence alignment of 13 acid proteases shown in fig. 1A and B.

search, multiple sequence alignment for homology search, construction of phylogenetic tree, common and conserved motif finding and protein-protein interaction. Physiochemical parameter analysis of individual protein sequences will help to understand the different physiochemical conditions for each individual protein which maintains their stability and also indicate respective organisms' optimum cultural conditions. Superfamily and phylogenetic tree analysis will help to classify the proteins and their evolutionary relatedness as well as protein-protein interaction analysis will detect the enzymes responsible for pathogenesis. Finally, consensus sequences from multiple sequence alignment and conserved motifs will help to design specific primers for each different species.

MATERIALS AND METHODS

Sequence retrieval

A total of 84 acid, 283 neutral and 31 alkaline proteases of *B. anthracis* origin were downloaded from NCBI (http://www.ncbi.nlm.nih.gov/) database. Among them 13 acid proteases, 9 neutral and one alkaline protease sequences were selected for *in silico* analysis.

Physiochemical parameters analysis

Physicochemical data were generated from ProtParam software

using ExPASy server (the proteomic server of Swiss Institute of Bioinformatics). FASTA sequence format were applied for subsequent analysis.

Protein superfamily and family search
The Superfamily tool on ExPASy server was used for protein family search.

Multiple sequence alignment (MSA)
The program ClustalW2 (Larkin et al., 2007) was used for multiple sequence alignment and MSA was represented by CLC-Bio sequence viewer.

Phylogenetic tree construction
Phylip-3.69 (Tuimala, 1989) was used for phylogram construction by Neighbor-joining (NJ) method using 100 bootstrap values. Tree was edited by Dendroscope (Huson et al., 2007).

Protein-protein interaction
Selected protein sequences were studied for protein-protein interaction to detect the probable function using STRING Database.

Motif finding
Acid and neutral proteases were separately subjected to Pfam to find out conserved domains. Separated domains were subjected to Block Maker for conserved block identification. Separated blocks were used for motif finding using MEME Suite. Conserved motif of alkaline protease was deduced from the multiple sequence alignment.

RESULTS

Among all the deposited *B. anthracis* protease sequences in the

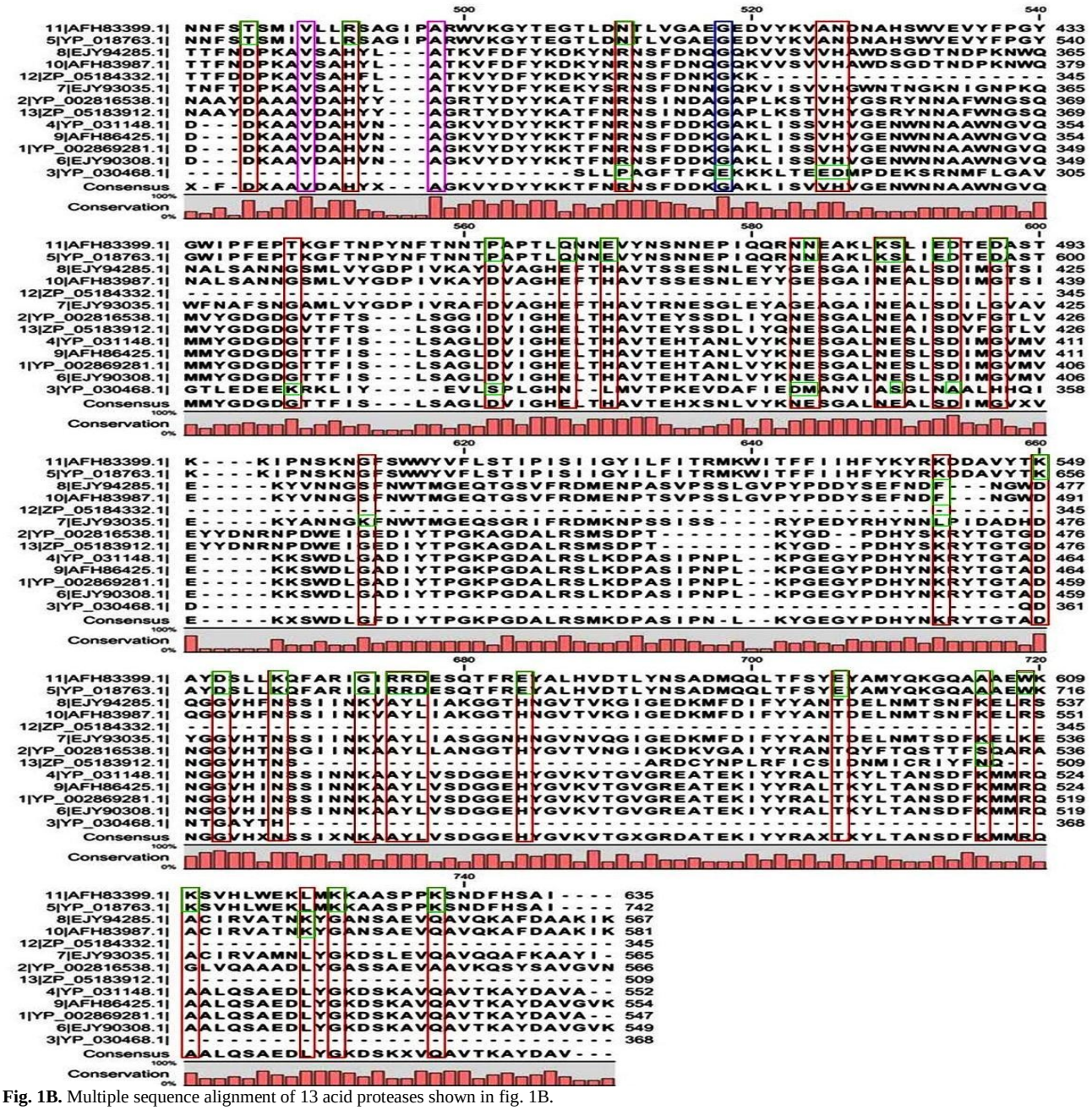

Fig. 1B. Multiple sequence alignment of 13 acid proteases shown in fig. 1B.

NCBI database, 13 acid, 9 neutral and one alkaline protease sequences were found to have unique sequences i.e. they all have sequence level dissimilarity and different amino acid compositions. The accession numbers of protease protein sequences along with the source organism and the type of proteases are listed in Table 1.

Physicochemical parameter analysis of proteases

The physicochemical features of protease sequences were represented in Table 2. The amino acid number for acid protease ranged from 345 to 742 with variable molecular weight. The pI value varied from 4.94 to 6.29, except sequences 5 and 11 (Acc. No. YP_018763.1 and AFH83399.1), which have their pI of 8.65 and 7.24 respectively. The above mentioned two sequences (5 and 11) were membrane bound proteases which have the aliphatic index value of 84.35 and 77.07 respectively. On the other hand sequence 3 (Acc. No. YP_030468.1) has the highest aliphatic index value of 96.36.

For all the neutral proteases group of protein different range were found in different analysis (Table 2). Germination protease (sequence 2) showed all the values similar to that of the germination protease of acid protease group (sequence3) with 4.94 pI value. Some serine proteases were also found in this group with various pI values. Accession number EJT19501.1 (Sequence 7) which was a membrane bound protease, showed the highest pI value of 9.33 and aliphatic index value of 104.64.

Protein superfamily and family search

The entire sequences of acid, neutral and alkaline proteases when subjected to Superfamily tools on ExPASy server revealed different superfamily and family (Table 3). For acid protease 10 sequences were found with Metalloproteases ("zincins"), catalytic domain superfamily and Thermolysin-like family (Table 3). Sequence 3 (Acc. No. YP_030468.1) was found with HybD-like superfamily and Germination protease

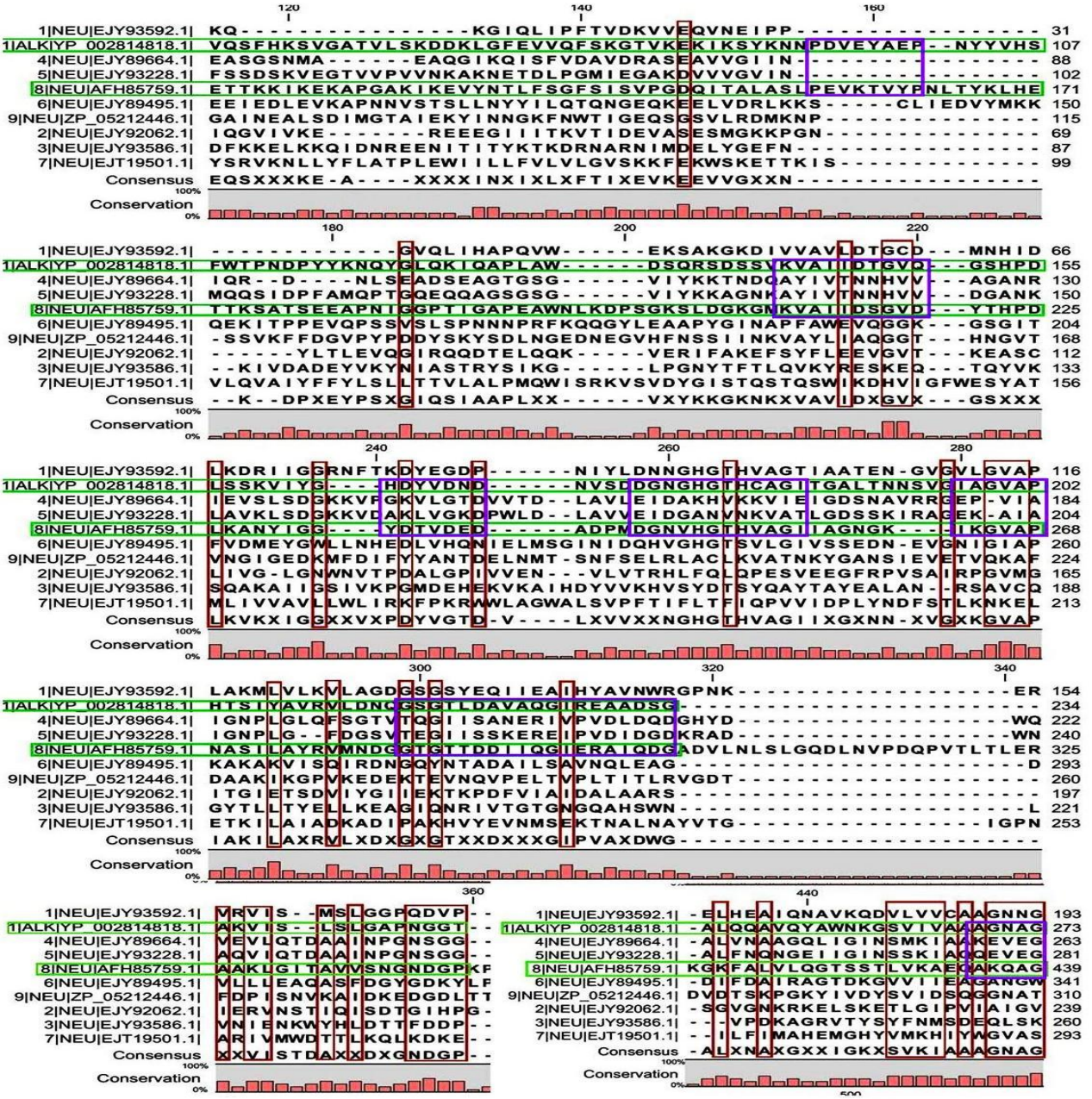

Fig. 1C. Fig. 1C represented multiple sequence alignment among 9 neutral proteases along with one alkaline proteases.

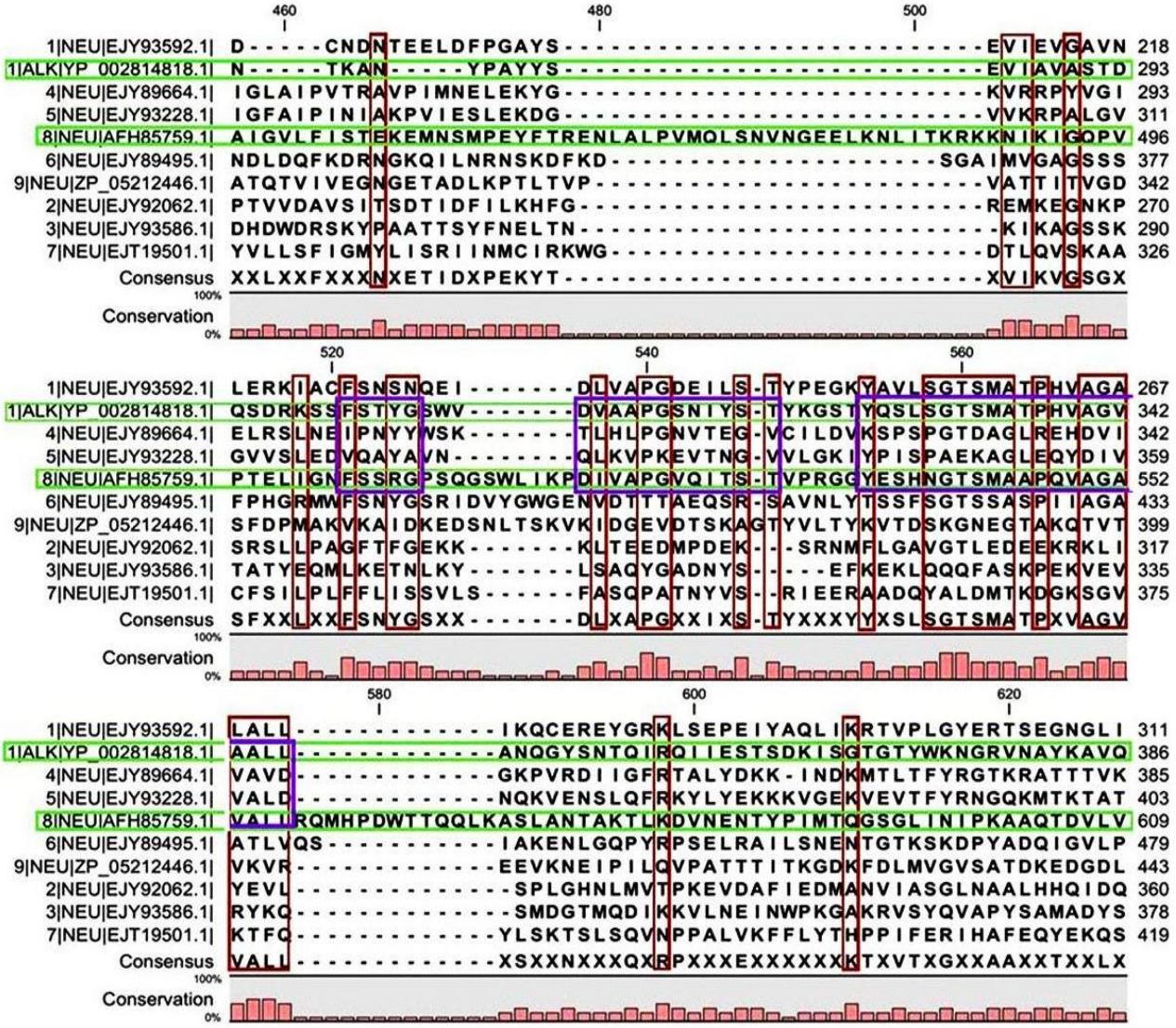

Fig. 1D. Fig. 1D represented multiple sequence alignment among 9 neutral proteases along with one alkaline proteases. Pink bars-highly conserved regions. Blue bars - conserved except sequence 3 (Acc. No. YP_030468.1). Red bars – near about conserved regions with change indicated in green bars. Red bar- highest similarity area. Horizontal green bar- similarity between alkaline protease (Acc. No.YP_002814818.1) and neutral protease sequence 8 (Acc. No. AFH85759.1). Ten violet box-short conserved regions of alkaline protease with respect to sequence 8.

family. Sequence 5 and 11 (Acc No. AFH83399.1 and YP_018763.1) were found with Cysteine proteinases superfamily and Transglutaminase core family. The short segments were found to have similarity with thermostable phytase (Table 3). But in the case of neutral proteases most variable domains were observed in superfamily and family analysis (Table 3). Among them sequence 7 and 9 were related to ten acid proteases, sequence 2 showed similarity with germination protease sequence 3 (acid protease) and sequence 8 specified similarity with alkaline protease 1. So a sequence level dissimilarity was observed among 9 neutral proteases. These have been reflected in the multiple sequence alignment also.

Multiple sequence alignment

Multiple sequence alignment analysis of the 13 acid proteases, the 9 neutral proteases and the one alkaline protease displayed the superfamily results of each groups. Fig. 1A and B showed consensus regions of acid proteases. Presence of consensus regions throughout the whole alignment indicated high level of sequence similarity among them. Three 100% conserved positions were found in aligned region such as position 368, 489 and 498 which have been represented in pink bar. Blue bars represented some specific changes only for sequence 3 (germination protease). Red bars represented near about conserved regions where green bars indicated the changes. In maximum cases, changes were found for two membranes bound proteases sequences 5 and 11 (Accession No. YP_018763.1 and AFH83399.1) and germination protease sequence 3 (Accession No. YP_030468.1).

A few ranges of consensus regions were found with low levels of sequence similarity in multiple sequence alignment of neutral protease group and one alkaline protease. Red bars indicated the similarity area. As the alkaline protease sequence (Accession No. YP_002814818.1) showed highest similarity with neutral protease sequence 8 (AFH85759.1) in previous experiments, they have been presented in green colour (Fig. 1C and D). Ten short conserved motifs of alkaline protease were also showed in violet box in comparison with neutral protease 8.

Phylogenetic tree construction

Phylogenetic tree construction of all the 13 acid protease, the 9

neutral protease and the one alkaline protease showed an interesting result. It was found that 10 acid protease were cluster together in the top of the tree (Fig. 2). Two membranes bound protease sequences 5 and 11 (YP_018763.1 and AFH83399.1) were found together in the bottom of the tree. Germination protease sequence 3 was found with another germination protease sequence 2 of neutral protease group. One alkaline serine protease was found with the neutral serine protease sequence 8.

Protein-protein interaction

Protein-protein interactions are the core of the interactom study which also represents the secretom of an organism. Here in this study the interaction of acid, neutral and alkaline protease of *B. anthracis* were studied. Among the entire category, only 9 acid proteases (Accession No. YP_002869281.1, YP_002816538.1, YP_031148.1, EJY90308.1, EJY93035.1c, EJY94285.1, AFH86425.1, AFH83987.1 and ZP_05183912.1) showed interaction with immune inhibitor A metalloproteases (inhA or others). Accession number YP_002869281.1 (NprB), EJY94285.1 (extracellular protease) and ZP_05183912.1 (Npr599) specifically showed interaction with inhA1 and AFH83987.1 interacts with inhA2. For two germination protease, interaction was found with some proteins affecting the sporulation and germination procedure of *B. anthracis*, like Putative stage II sporulation protein P, spore cortex-lytic enzyme prepeptide, germination protein YpeB, Stage IV sporulation protein A, Small acid-soluble spore protein/B etc.

Motif finding

A total of 6 motifs were found from acid and neutral protease (Table 4). Motif A2 and A3 showed similarity with peptidase M4 function as per the BLAST and PFAM result. According to PFAM and GENE3D motif B1, 2, and 3 all have the peptidase activity. The function of B3 deduced by BLAST was endopeptidase spore protease Gpr. Ten short conserved motifs of alkaline protease were identified from multiple sequence alignment (Fig. 1C and D).

DISCUSSION

The present study reported that *B. anthracis* acid metalloproteases have some definite role in their pathogenesis. The extracellular nature and the protein-protein interaction pattern claimed the involvement of some acid proteases during anthrax infection.

Physicochemical nature of a protein can be easily calculated through *in silico* analysis based on their amino acid sequences. Solubility of protein can be determined from the Grand average of hydropathicity or GRAVY. Positive GRAVY indicates hydrophobicity and negative GRAVY indicates the hydrophilicity (Kyte, 1982). On the other hand thermostability of a protein is directly proportional to the aliphatic index value.

All the metalloproteases of acid group were found to prefer

Table 3. Distribution of superfamily and family among acid, neutral and alkaline proteases of *Bacillus anthracis*

Superfamily	Family	No.	Accession number(range of amino acid residue)	Total No.
Acid				
HybD-like	Germination protease	3	|YP_030468.1|(3 - 368)	1
Metalloproteases ("zincins"), catalytic domain	Thermolysin-like	1	|YP_002869281.1|(225 - 546)	10
		2	|YP_002816538.1|(251 - 566)	
		4	|YP_031148.1|(230 - 551)	
		6	|EJY90308.1|(225 - 549)	
		7	|EJY93035.1|(244 - 562)	
		8	|EJY94285.1|(246 - 566)	
		9	|AFH86425.1|(230 - 554)	
		10	|AFH83987.1|(260 - 580)	
		12	|ZP_05184332.1|(247 - 344)	
		13	|ZP_05183912.1|(255 - 484)	
Cysteine proteinases	Transglutaminase core	11	|AFH83399.1|(245 - 557)	2
		5	|YP_018763.1|(352 - 664)	
Thermostable phytase (3-phytase)	Thermostable phytase (3-phytase)	11a	|AFH83399.1|(155 - 273)	
		5a	|YP_018763.1|(263 - 380)	
Neutral				
HybD-like	Germination protease	2	|EJY92062.1|(3 - 368)	1
Subtilisin-like	Subtilases	1	|EJY93592.1|(30 - 314)	
		6	|EJY89495.1|(180 - 463)	
		8	|AFH85759.1|(184 - 368, 502 - 604)	
Collagen-binding domain	Collagen-binding domain	6a	|EJY89495.1|(503 - 612)	
Protease propeptides/inhibitors	Subtilase propeptides/inhibitors	8a	|AFH85759.1|(124 - 168)	
PA domain	PA domain	8b	|AFH85759.1|(394 - 460)*	
Cysteine proteinases	Transglutaminase core	3	|EJY93586.1|(132 - 273)	1
Trypsin-like serine proteases	Prokaryotic proteases	4	|EJY89664.1|(69 - 286)	2
		5	|EJY93228.1|(84 - 304)	
PDZ domain-like	HtrA-like serine proteases	4a	|EJY89664.1|(271 - 386)	
		5a	|EJY93228.1|(288 - 403)	
Metalloproteases (“zincins”), catalytic domain	Matrix metalloproteases, catalytic domain	7	|EJT19501.1|(220 - 286)	1
Metalloproteases ("zincins"), catalytic domain	Thermolysin-like	9	|ZP_05212446.1|(20 - 229)	1
LDH C-terminal domain-like	AglA-like glucosidase	9a	|ZP_05212446.1|(267 - 339)	
PKD domain	PKD domain	9b	|ZP_05212446.1|(267 - 339)*	
Alkaline				
Subtilisin-like	Subtilases	1	|YP_002814818.1|(119 - 390)	1
Protease propeptides/inhibitors	Subtilase propeptidae/ inhibitors	1a	|YP_002814818.1|(37 - 106)	

extracellular medium according to GRAVY results (Table 2) and they were moderately thermostable in nature. In reference to the above parameters protein sequences of 5 and 11 were highly thermostable and membrane bound proteases of *B. anthracis*. The presence of thermostable phytase (3-phytase) domain in sequence 5 and 11 at superfamily and family analysis also supported their thermostability (Table 3). Germination protease (sequence 3) showed highest aliphatic index value of 96.36 indicating high level of thermal stability. The 4.94 pI value of two germination protease indicated that the germination of *B. anthracis* occurs in acidic medium and the high value of hydrophobicity (GRAVY -0.235) denied their extracellular existence (Table 2). In case of neutral protease group sequence 7 was found to prefer intracellular medium with highest thermal stability of 104.64. Sequence 3 and 6 showed hydrophilic nature but all other were hydrophobic or preferred intercellular space. The in vivo half-life of a protein can be calculated in the form of instability index (Guruprasad et al., 1990). As per the literature, proteins having more than 40 instability index value, having less than 5h of half-life and proteins having less than 40 instability index values, having more than 16h of half-life (Rogers et al., 1986). From this point of view the studied proteases have their half life of greater than 16 h.

Phylogenetic tree (Fig. 2) visibly reflected the superfamily results (Table 3). Although all the proteins were sequentially different, 10 acid proteases showed close relationship according to their evolution. They all showed same domains in superfamily search indicating their similar function. Functional similarity was also found between two germination proteases and also between two membrane bound proteases. The studied neutral proteases were found together in the tree. Among them neutral serine protease 8 showed highest similarity with alkaline serine protease one (Fig. 2), indicating their functional similarity. From the above result it was found that same sequence represented same functional domain and on the basis of that they showed evolutionary relationships.

According to Baillie (2001) and Russell et al. (2007) anthrax pathogen initiates their germination as well as infection by interacting with the host macrophages. The resulting vegetative cells spread in blood and other tissues as the causative agent of meningitis and ultimately causes death. Literature showed that among the secreted metalloproteases of *B. anthracis*, immune inhibitor A1 (InhA1) was found to be the single pathogenic member and during the infection it helps to cleave mammalian cell matter with addition to the modulation of the *B. anthracis* secreted proteins (Pflughoeft et al., 2014) and increases the permeability of blood-brain barrier resulting cerebral hemorrhages (Dhritiman et al., 2011). Previous investigation of Pflughoeft (2010) suggested that the protease cascade regulated the organism response in altering environments like reacting to a changing signal or in presence of different types of tissue. Mukherjee et al. (2011), also showed that NprB and Npr599 are the extracellular enzymes which interact with inhA1 and were able to degrade the plasma and matrix proteins of host. From the secretom analysis through protein-protein interaction, it has been shown that maximum acid proteases (except 2, 3, 5, 6, and 11) interacts with immune inhibitor A (inhA or othrs) as shown in Table 5. So, in accordance to the above literature it can be concluded that protease from the organism with accession number YP_002869281.1, YP_002816538.1, YP_031148.1, EJY90308.1, EJY93035.1, EJY94285.1, AFH86425.1, AFH83987.1, ZP_05184332.1 and ZP_05183912.1 (Table 1) are extracellular in nature and they are proved to be the extracellular protein in this study through physicochemical parameter analysis. Among them zinc metalloprotease (Table 3) YP_002869281.1 (NprB), EJY94285.1 (extracellular protease) and ZP_05183912.1 (Npr599) are the part of regulatory cascade of *B. anthracis* which helps the cell to react against changed external environment and supports its stabilization in altered condition and may have some direct relation to the permeability of endothelial cell and degradation of plasma or matrix protein at the time of infection (Mukherjee et al., 2011). Thus the above evidence linked physicochemical parameters and PPI outputs as a result of which extracellular and

Table 4. Identified motifs for acid and neutral proteases with their function deduced by protein BLAST and INTERPROSCAN. Accession number, query coverage, e-value and maximum identity of highly similar sequence are represented here

Motif No.	Motif Width	Sequence	Occurrence in keratinase sequences	Function deduced by BLAST	Acc. No. [Name] of Similar sequence from BLAST	Query coverage	E value	Max identity	Function deduced by INTERPROSCAN
Acid									
A1	29	ESGTTHVRLQQKHNNI PVYGSDQTVHLDK	10	Neutral protease B	WP_002058328.1 [Bacillus cereus]	100%	2e-15	90%	FTP domain (PFAM)
A2	50	DNGGVHINSSIINKAA YLISDGGEHYGVKVT GIGREKMFKIYYRALT KYL	8	peptidase M4	WP_000745148.1 [Bacillus cereus]	100%	1e-20	88%	Peptidase M4, C-terminal(PFAM)
A3	41	AAVDAHYLAGKVYDY YKKTFNRNSFDDQGQ KLISSVHCWER	8	peptidase M4	WP_000344172.1 [Bacillus thuringiensis]	97%	1e-13	83%	Peptidase M4 Domain (PFAM)
Neutral									
B1	49	KHVYEVNMSEKTNAL NAYVTGIGPNARIVM WDTTLKQLKDKEILFI MAH	1	peptidase M48	WP_002020579.1 [Bacillus cereus]	100%	5e-26	100%	Peptidase M48 (PFAM)
B2	49	MAAPQVAGAVALLRQ MHPDWTTQQLKASLA NTAKTLKDVNENTYPI MTQ	2	minor extracellular protease VpR	NP_846805.1 [Bacillus anthracis str. Ames]	100%	3e-24	100%	Peptidase S8/S53 domain (GENE3D)
B3	50	MAAPQVAGAVALLRQ MHPDWTTQQLKASLA NTAKTLKDVNENTYPI MTQ	1	Endopeptidase spore protease Gpr	YP_005120931.1 [Bacillus cereus F837/76]	100%	1e-26	100%	Peptidase S8/S53 domain (GENE3D)

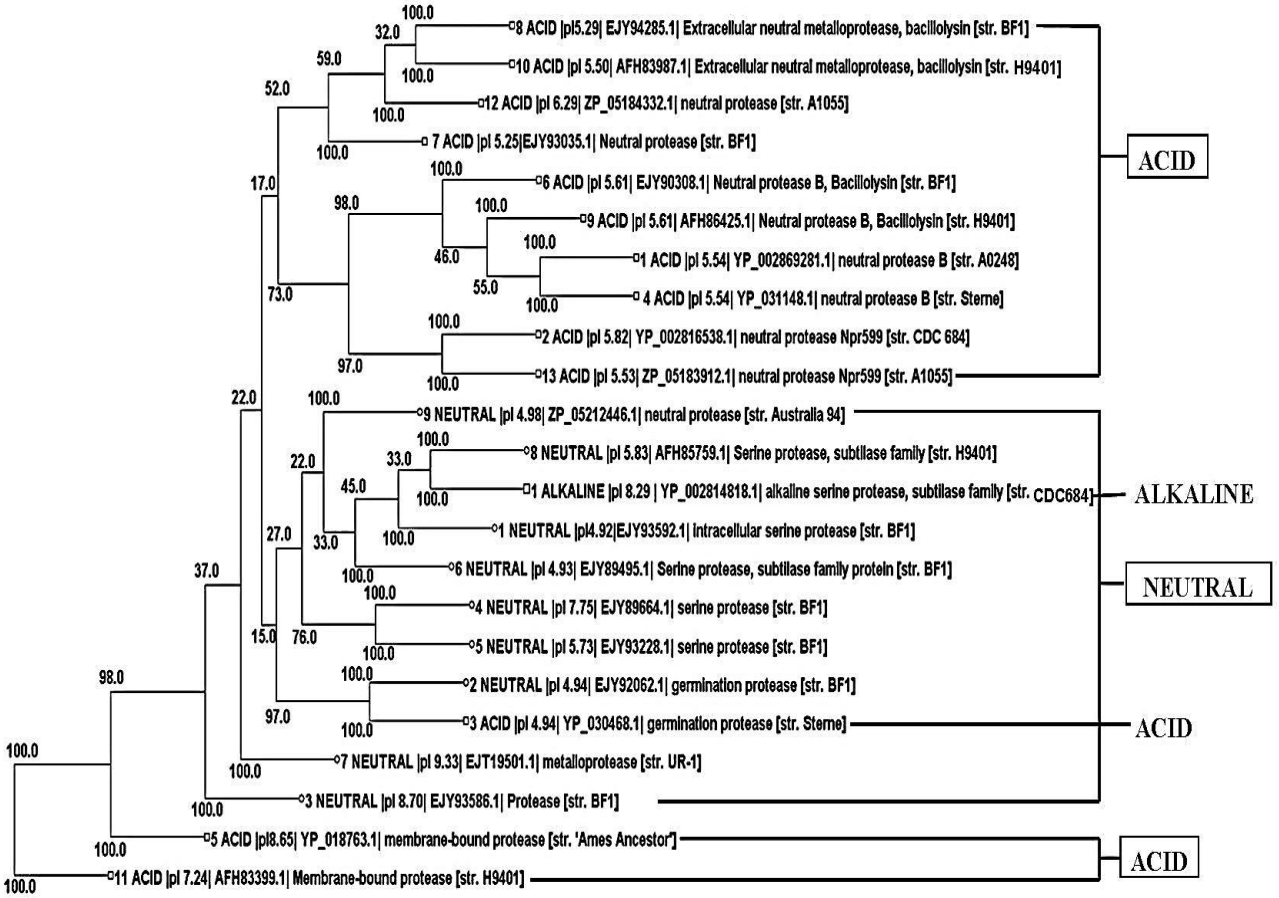

Fig. 2. Phylogram of 13 acid protease, 9 neutral proteases and 1 alkaline proteases. Acid and neutral proteases were separated in two distinct groups. Except acid germination protease (Accession no. YP_030468.1), showed similarity with neutral germination protease. Two membrane bound acid protease (Accession no. YP_018763.1 & AFH83399.1) were also found together in the bottom of the tree. Alkaline protease was found with neutral protease sequence 8 (Accession no. AFH85759.1) according to its serine protease property. Phylogram was also indicating the evolutionary changes.

intracellular acid proteases could be identified. As neutral proteases have diversities in their sequence, they showed different interactions also (Table 5). Further investigation on neutral and alkaline protease PPI study is needed for the better understanding of their biological functions.

As 100% sequence similarity were found among 31 retrieved alkaline proteases, it can be concluded that *B. anthracis* alkaline protease are more conserved than others. A total of 10 aligned regions were found for alkaline protease 1 during the multiple sequence alignment result analysis in comparison with neutral protease sequence 8 (Fig. 1C and D). Identified short segments could be used as primer or probe to identify *B. anthracis* alkaline protease. Besides specific regions in multiple sequence alignment results of 13 acid proteases (Fig. 1A and B) could be used for further investigation. The pink bars in acid protease group were the highly conserved regions which could be used as the target site for the inactivation of those proteases. Acid and neutral protease specific motifs were represented in Table 4 where peptidase activity was found for A2, A3, and B1. Similarity was found between B3 and *B. cereus* endopeptidase spore protease Gpr, which could be related with the germination procedure of *B. anthracis* spore. For the preparation of acid protease specific primers and probes, or to inactivate the protease responsible for spore germination, identified motifs could be used. The detailed study of metabolic network could be investigated further.

CONCLUSION

The *in silico* characterization of *B. anthracis* protease revealed pH range based sequence similarity. Multiple sequence alignment and motif finding result can be used to design degenerate primers or probes for specific sequences as to cloning the putative genes based on PCR amplification for further analysis. Conserved amino acid positions could be used as target site to deactivate the enzyme function. Among all the groups only acid protease were found to interact with InhA, which indicated that metallo proteases of acid protease group have the capability to develop pathogenesis during *B. anthracis* infection. Deactivation of conserved amino acid position of germination protease can stop the sporulation and germination of *B anthracis* cell. The detailed interaction study of neutral and alkaline proteases could also help to design the interaction network for the better understanding of anthrax disease. Further study on structure prediction and protein-protein or protein-ligand interaction of *B. anthracis* proteases could reveal new drugs to inactivate the disease causing proteins.

ACKNOWLEDGEMENTS

Authors are very much thankful to UGC, Govt. of India for providing fellowship to the Department of Microbiology, Vidyasagar University and also the Department of Biotechnology, Govt. of India is acknowledged gratefully for creation of BIF center in Vidyasagar University.

Table 5. List of protein-protein interaction study. Accession number, name and functions of interacted proteins were listed here

Acid				
Type of Protease	**Protein No.**	**Accession No. of Protein (PPI protein)**	**Interacted Partner**	**Function of the Partner**
Neutral protease B	1	YP_002869281.1	BAS4908	Hypothetical protein
	4	YP_031148.1	BAS1197	Immune inhibitor A metalloprotease InhA1
	6 (Bacillolysin)	EJY90308.1 (except inhA & BAS2003)	inhA	Immune inhibitor A metalloprotease
			BAS2028	Putative neutral metalloprotease
	9 (Bacillolysin)	AFH86425.1	BAS2003	Hypothetical protein
			BAS4909	MarR family transcriptional regulator
Neutral protease Npr599	2	YP_002816538.1 (BAMEG_3914)	BAS1197	Immune inhibitor A
	13	ZP_05183912.1 (inhA)	BAMEG_3914 / inhA	Immune inhibitor A / Immune inhibitor A metalloprotease
			BAS0566	Transcriptional regulator/TPR domain protein
			BA_0598	Hypothetical protein
			BAS2028	Putative neutral metalloprotease
Germination protease	3	YP_030468.1	BAS4219	Hypothetical protein
			spoIIP	Stage II sporulation protein P
			BAS2886	Putative stage II sporulation protein P
			sleB	Spore cortex-lytic enzyme prepeptide
			spoIVA	Stage IV sporulation protein A
			ypeB	Germination protein YpeB
			sasP1	Acid-soluble spore protein
			sspB	Acid-soluble spore protein B
			sasP2	Acid-soluble spore protein
			BAS2908	Acid-soluble spore protein, alpha/beta family
Membrane-bound protease	5	YP_018763.1	BAS1973	Hypothetical protein
	11	AFH83399.1 (except BAS1975)	BAS1974	Hypothetical protein
			BAS2824	AAA family ATPase
			BAS2803	Hypothetical protein
			BAS4373	Hypothetical protein
			BAS2804	Hypothetical protein
			BAS1422	Putative heptaprenyl diphosphate synthase component I
			BAS1975	Hypothetical protein
Neutral protease	7	EJY93035.1c	BAS2532	Hypothetical protein
			BAS3190	Enhancin family metallprotease
			BAS1197	Immune inhibitor A
			BAS2028	Putative neutral metalloprotease
			BAMEG_3914	Immune inhibitor A
	12	ZP_05184332.1	BCAH820_2580	Hypothetical protein
			BCAH820_2582	Hypothetical protein
			BCAH820_0417	Phage infection protein
			BCAH820_2212	Group-specific protein
Extracellular neutral metallo- protease, bacillolysin	8	EJY94285.1	BCAH820_1373	Immune inhibitor A metalloprotease InhA1
			inhA	Immune inhibitor A metalloprotease
			BCAH820_2731	Hypothetical protein
			BCAH820_2202	Putative neutral metalloprotease
	10	AFH83987.1	BALH_1145	Immune inhibitor A
			inhA2	Immune inhibitor A
			BALH_2445	Hypothetical protein
			BALH_1941	Bacillolysin
Neutral				
Type of Protease	**Protein No.**	**Accession No. of Protein**	**Interacted Partner**	**Function of the Partner**
intracellular serine protease	1	EJY93592.1	BAS1858	Putative transition state transcriptional regulatory protein
			glnA	Type I glutamine synthetase
germination protease	2	EJY92062.1	BAS4219	Hypothetical protein
			spoIIP	stage II sporulation protein P
			BAS2886	Putative stage II sporulation protein P
			sleB	Spore cortex-lytic enzyme prepeptide
			spoIVA	Stage IV sporulation protein A
			ypeB	Germination protein YpeB
			sasP1	Small acid-soluble spore protein
			sspB	Small acid-soluble spore protein B
			sasP2	Small acid-soluble spore protein
			BAS2908	Alpha/beta family small acid-soluble spore protein
Protease	3	EJY93586.1	BAS0846	Hypothetical protein
			BA_1245	Hypothetical protein
			BAS1853	Hypothetical protein
			BAS2989	Putative penicillin-binding protein

Table 5. List of protein-protein interaction study. Accession number, name and functions of interacted proteins were listed here (continued)

Protease	3	EJY93586.1	BAS3973	Hypothetical protein
			BAS1851	Hypothetical protein
			BAMEG_2594	Hypothetical protein
			BAS1854	MutT/NUDIX family protein
			BAS4125	ATP-binding protein ABC transporter
			gerXB	Allow B.anthracis to germinate within phagocytic cells and therefore involved in virulence
Serine protease	4	EJY89664.1	rlmH	rRNA large subunit methyltransferase that methylates the pseudouridine at position 1915 (m3Psi1915) in 23S rRNA
			ftsH	Cell division protein FtsH
			groL	Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions
			BAS5315	Metallo-beta-lactamase family protein
			dnaK	Chaperone protein DnaK acts as a chaperone
			BAS5032	Metallo-beta-lactamase family protein
			BAS5316	YycI protein
			yycG	Sensory box histidine kinase YycG
			BAS5313	Hypothetical protein
			yycF	DNA-binding response regulator YycF
Serine protease	5	EJY93228.1	groL	Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions
			dnaK	Chaperone protein DnaK acts as a chaperone
			ftsH	Cell division protein FtsH
			clpX	ATP-dependent protease ATP-binding subunit ClpX directs the protease to specific substrates
			BAS5032	Carboxyl-terminal protease
			sigA	RNA polymerase sigma-43 factor promotes the attachment of RNA polymerase to specific initiation sites and are then released
			grpE	Participates actively in the response to hyperosmotic and heat shock by preventing the aggregation of stress-denatured proteins in association with dnaK and grpE. It is the nucleotide exchange factor for dnaK and may function as a thermosensor
			clpP2	ATP-dependent Clp protease with proteolytic subunit ClpP cleaves peptides in various proteins in a process that requires ATP hydrolysis. It has a chymotrypsin-like activity. It plays a major role in the degradation of misfolded proteins
			aroA	3-phosphoshikimate 1-carboxyvinyltransferase
			groS	Act as chaperonin and binds to Cpn60 in the presence of Mg-ATP and suppresses the ATPase activity of the latter
Serine protease, subtilase family protein	6	EJY89495.1	vpR	Minor extracellular protease VpR
			BAS2216	Subtilase family alkaline serine protease
Metallprotease	7	EJT19501.1	gcp	putative O-sialoglycoprotein endopeptidase
			lepA	GTP-binding protein LepA
			BAS1044	Hypothetical protein
			BAS1046	Methyl-accepting chemotaxis protein
			hemK	Protein-(glutamine-N5) methyltransferase
			rsgA	Ribosome-associated GTPase; May play a role in 30S ribosomal subunit biogenesis. Unusual circulary permuted GTPase that catalyzes rapid hydrolysis of GTP with a slow catalytic turnover
			BAS1043	Hypothetical protein
			nth	Endonuclease III
			dnaK	Chaperone protein DnaK
			ftsZ	Cell division protein FtsZ; essential in the cell-division process; binds to and hydrolyzes GTP
Serine protease, subtilase family	8	AFH85759.1	BCAH820_2580	Hypothetical protein
			BCAH820_2582	Hypothetical protein
			BCAH820_0417	Phage infection protein
			BCAH820_2212	Group-specific protein
Neutral protease	9	ZP_05212446.1	BAS3605	Subtilase family Serine protease
			BAA_4604	Hypothetical protein
			BAS4251	TolB domain protein
			rlmH	rRNA large subunit methyltransferase, specifically methylates the pseudouridine at position 191
			BAS5208	Putative aminopeptidase

Table 5. List of protein-protein interaction study. Accession number, name and functions of interacted proteins were listed here (continued)

Alkaline				
Alkaline serine protease, subyilase family	1	YP_002814818.1	glnA	type I glutamine synthetase
			BAS2217	Sodium/alanine symporter family protein
			hemE	Catalyzes the decarboxylation of four acetate groups of uroporphyrinogen-III to yield coproporphyrinogen-III
			BAS3605	Subtilase family serine protease

CONFLICT OF INTEREST

The authors declare that there was no conflict of interest.

REFERENCES

Baillie L. The development of new vaccines against *Bacillus anthracis*. J Appl Microbiol. 2001;91:609-613.

Banerjee A, Jana A, Pati BR, Mondal KC, Das Mohapatra PK. Characterization of tannase protein sequences of bacteria and fungi: an *in silico* study. Protein J. 2012;31:306-327.

Chertow JH. *Bacillus anthracis* protease regulates bacterial adhesion. ProQuest Dissertations and Theses of George Mason University. 2011.

Dubey AK, Yadav S, Kumar M, Singh VK, Sarangi BK, Yadav D. *In silico* characterization of pectate lyase protein sequences from different source organisms. Enzyme Res. 2010;2010:1-11.

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;4:155-161.

Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinform. 2007;8:460-464.

John WE, Teresa GA. Serum protease cleavage of *Bacillus anthracis* protective antigen. J Biol Chem. 2010;285:8130-8137.

Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157:105-132.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, William HM, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and clustal X version 2.0. Bioinformatics. 2007;23:2947-2948.

Morya VK, Yadav S, Kim EK, Yadav D. *In silico* characterization of alkaline proteases from different species of *Aspergillus*. Appl Biochem Biotechnol. 2012;166:243-257.

Mukherjee DV, Tonry JH, Kim KS, Ramarao N, Popova TG, Bailey C, Mail SP, Chung MC. *Bacillus anthracis* protease inha increases blood-brain barrier permeability and contributes to cerebral hemorrhages. PLoS ONE. 2011;6:1-11.

Pflughoeft KJ. The Immune Inhibitor A1 Protease of *Bacillus anthracis*. Dissertations and theses of graduate School of Biomedical Sciences in the University of Texas. 2010.

Pflughoeft KJ, Swick MC, Engler DA, Yeo HJ, Koehler TM. Modulation of the *Bacillus anthracis* secretome by the immune inhibitor a1 protease. J Bacteriol. 2014;196:424-435.

Pomerantsev AP, Pomerantseva OM, Moayeri M, Fattah R, Tallant C, Leppla SH. A *Bacillus anthracis* strain deleted for six proteases serves as an effective host for production of recombinant proteins. Protein Expr Purif. 2011;80:80-90.

Rogers S, Wells R, Rechsteiner M. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science.1986;234:364-368.

Russell BH, Vasan R, Keene DR, Xu Y. *Bacillus anthracis* internalization by human fibroblasts and epithelial cells. Cell Microbiol. 2007;9:1262-1274.

Sheppard MR, Vinay K, Abbas AK, Nelson F. Robbins Basic Pathology. 8th ed. (Philadelphia, USA: Saunders), pp. 1-122, 2007.

Shrivastava S, Shukla P, Poddar R. *In silico* studies for evaluating conservation homology among family 11 xylanases from *Thermomyces lanuginosus*. JASES. 2007;2:70-76.

Tonry JH, McNichol BA, Ramarao N, Chertow DS, Kim KS, Stibitz S, Schneewind O, Kashanchi F, Bailey CL, Popov S, Chung MC. *Bacillus anthracis* protease InhA regulates BslA-mediated adhesion in human endothelial cells. Cell Microbiol. 2012;14:1219-1230.

Tuimala J. A primer to phylogenetic analysis using the PHYLIP package. Cladistics. 1989; 5:164-166.