Chelonian Conservation And Biology

Vol. 18 No. 2 (2023) | https://www.acgpublishing.com/ | ISSN - 1071-8443

DOI: doi.org/10.18011/2023.11(2).669.688

CHARACTERIZATION OF ALKALINE PROTEASE FROM BACILLUS CEREUS STRAIN S8 BY INSILICO APPROACH

Lakshmi B.K.M[1*], Muthyam Sowjanya[2], Muni kumar D[3], Ch.Venkatrayulu[4] and Hemalatha H.P.J[5]

[1,3,5] Department of Biochemistry, Andhra University, Visakhapatnam, Andhra Pradesh, India-530003.

[2, 4] Department of Marine Biology, Vikrama Simhapuri University, Kakutur, Nellore, Andhra Pradesh, India-524324.

***Corresponding Author:** kamalab42@gmail.com

Abstract

Proteases are the hydrolytic enzymes which catalyzes hydrolysis of proteins and one of the most prominent group of enzymes that find potential uses in various industrial sectors. Scientists have been trying continuously to introduce the enzymatic methods instead of chemical processes to promote the ecofriendly environment. The present study aims to carry such superior alkaline protease (APBCS8) insilico characterization from *Bacillus cereus* strain S8 (APBCS8) as deducing its unique three dimensional (3D) native structure is a daunting task. It contains 208 amino acids. Conserved domain, motif scan results and phylogenetic tree analysis of APBCS8 revealed that it belongs to Beta-lactamase family and transpeptidase super family. Three dimensional structure of APBCS8 was predicted by using LOMETS, a threading based structure prediction method. The modeled secondary structure of protease contains 3 beta sheets, 4 beta hairpins, 1 beta bulge, 9 strands, 10 helices, 9 helix helix interacs, 17 beta turns and 2 gamma turns. Assessment and authentication results from Ramachandran plot analysis by PROCHECK showed that alkaline protease from *Bacillus cereus* strain S8 (APBCS8) is of good quality. Active site of APBCS8 was predicted using COACH to predict ligand binding sites using the BioLiP database and demonstrated that this protease is a serine protease, probably mediates its mechanism of action through catalytic triad (Histidine, Aspartic acid, Serine). Interaction of APBCS8 with set of proteins is studied by using STRING database. It can be concluded that structural characterization of enzymes by bioinformatic tools might be fruitful in predicting their novel properties by studying at their gene level.

Key Words: Alkaline protease APBCS8, LOMETS, Three dimensional structure, COACH, STRING.

Introduction

Proteolytic enzymes play a specific catalytic role in the hydrolysis of proteins. Proteases are the most important industrial enzymes that perform a wide variety of functions and have various important biotechnological applications (Gupta *et al*., 2002). Among the various proteases, alkaline proteases are highly stable and active under harsh conditions. Alkaline proteases are of special interest as they are used in leather processing and manufacturing of detergents, food, pharmaceuticals (Dias *et al*., 2008).

Studies on structural analyses of alkaline proteases are limited and hence determining the 3D structures is of particular importance to provide insights into the structural principles that control the variety of enzyme properties and further conduct the molecular direct evolution (Jaouadi *et al.,* 2008).

Once a protein sequence has been determined, deducing its unique three dimensional (3D) native structure is a daunting task. Experimental methods to determine detailed protein structure, such as X ray diffraction studies and nuclear magnetic resonance (NMR) analyses, are highly labour intensive and the gap in the number between sequenced proteins and known structures increases. Thus, model building of proteins is of great importance. When a protein is first unfolded *in vitro* and then released again, it folds back to the same three dimensional structure it had before. Thus, various prediction methods are based on the assumption, that the three dimensional protein structure is determined by its primary structure. Structure prediction methods are coarsely divided into three categories: Comparative modelling, Fold recognition and *Ab initio* prediction.

The most reliable results are obtained, if the structure of a very close related sequence is known and comparative modelling can be applied. As the target protease sequence from the *Bacillus cereus* strain S8 exhibit greater similarity with the beta lactamase family, structural determination of protease was performed by homology modeling using LOMETS.

Proteins perform the biological functions through interactions with other molecules (called ligands). The identification of specific ligand-binding site (LBS) on proteins is often the first important step towards understanding the function of protein molecules, (Lopez, 2011, Roy and Zhang, 2012).

Prediction of protein-protein interactions is a field of combining bioinformatics and structural biology in order to identify and catalog physical interactions between pairs or groups of proteins. Knowledge of protein-protein interactions is important for the investigation of intracellular signaling pathways, modeling of protein complex structures and for gaining insights into various biochemical processes.

The present study describes the purification, biochemical and molecular characterization, kinetic studies as well as *insilico* analysis of extracellular alkaline protease from *Bacillus cereus*

strain S8 with the aim of understanding the properties of the enzyme and assessing its worthiness as a commercial enzyme.

Materials and Methods

Source of the organism

Bacillus cereus strain S8 (MTCC NO 11901), which was isolated from the soil of Upputeru lake, Kakinada by Lakshmi *et al*., (2014) was used for the production of crude alkaline protease.

Production of alkaline protease

Production of crude alkaline protease from *Bacillus cereus* strain S8 (MTCC NO 11901) was carried out by statistically optimized media formulated by (Lakshmi and Hemalatha, 2016). The protease was purified by chromatographic techniques which is optically active at pH 10.0 and temperature 70^{o}C with great stability towards pH and temperature with casein as a specific substrate (Lakshmi *et al*., 2018). Sequencing of protease by MALDI-TOF analysis showed that it is made up of 208 aminoacids with molecular weight of 23.3 kDa and this protease was subjected for *insilico* analysis for determination of conserved motifs, phylogenetic analysis, its interaction with other proteins in a cell as well as its structural and active site prediction by various bioinformatics tools as follows

Phylogenetic Analysis

The protease sequence of *Bacillus cereus* strain S8 was compared against the different types of protease sequences available from protein database using the p-BLAST program and they were aligned using CLUSTAL W software (Thompson *et al.*, 1994). The evolutionary history was inferred using the Minimum Evolution method (Rzhetsky and Nei, 1992). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al*., 2004). The Minimum Evolution (ME) tree was searched using the Close-Neighbor-Interchange (CNI) algorithm (Nei and Kumar, 2000). The Neighbor-joining algorithm (Saitou and Nei, 1987) was used to generate the initial tree. The MEGA 4 package (Tamura *et al.*, 2007) was used for all analyses.

Conserved Domain Database (CDD)

The conserved domain database (CDD) is a protein annotation resource that consists of a collection of well-annotated multiple sequence alignment models for ancient domains and full-length proteins.

Motif Scan

Motif scanning is a tool that helps in finding all known motifs that occur in a sequence and deals with the interpretation of the match scores (Sigrist, 2002). The matched sequences reported by search programs can be classified as true positives and false positives. Statistical analysis based on sound principles help in the decision because some matches are more likely to have been produced by chance than others.

Protein modeling by LOMETS

LOMETS generates 200 models for a given target which was generated by 10 component servers where each server generates 20 models as sorted by their Z-scores in each algorithm. The best 10 models are finally selected from the 160 models based on the following scoring function:

$$\textbf{Score (i,j)=Z(i,j)/Z0(i)*conf(i)+seqid (i,j)}$$

Where

Z(i,j) is the Z-score of jth model of the ith server

Z0(i) is the cutoff of ith server

Conf(i) is the confidence of ith server

Seqid (i,j) is the sequence identity to query of jth model of ith server

Model Evaluation using PDB Sum Generate

Model of query sequence that has been generated by LOMETS was evaluated using PDB Sum database which was provided by European Bioinformatics Institute. The PDB Sum generates PROCHECK results which consist of Ramachandran plot, secondary structure prediction, and topology information as well as information about clefts.

Prediction of active site of alkaline protease

COACH

COACH (**C**onsensus **A**pproa**CH**) is a metaserver approach to protein ligand binding site prediction. Starting from given structure of target proteins, COACH generates complementary ligand binding site predictions using two comparative methods, TM-SITE and S-SITE, which recognize ligand binding templates from the BioLip protein function database by binding specific substructure and sequence profile comparisons. These predictions will be combined with results from other methods (COFACTOR, FINDSITE and ConCavity) to generate final ligand binding site predictions.

Protein protein interactions

STRING

Many of the protein-protein associations in STRING are imported from other databases but STRING also contains a large body of predicted associations that are produced *de novo*. These predictions are based on systematic genome comparisons (Huynen *et al*., 2003). STRING assigned a confidence score to each predicted association based on the performance of the predictions against a common reference set of trusted and true associations. Further choose the functional grouping of proteins maintained at KEGG [Kyoto Encyclopedia of Genes and Genomes, (Kanehisa *et al*., 20004)] as the reference.

Results and Discussion

Peptide mass finger printing (PMF) was performed for purified alkaline protease from *Bacillus cereus* strain S8. It showed partial homology with hypothetical protein from *Bacillus* sp. **BLA3_BACCE** (Accession number- P06548.1) in the PMF database which is having superfamily of transpeptidase and family of beta-lactamase domain. The putative amino acid sequence was showing 66% sequence homology to that of the hypothetical protein from *Bacillus* sp. **BLA3_BACCE**. The sequences were aligned using the NCBI Blast research. The total number of amino acids 208 and the isoelectric point was found to be 9.52.

FTNSHYKKSNKQTNQTNQVKQENKRNHAFAKLEKEYNAKTVAYHADDRAVGALL RQNSIEALDERITYTRKAFASTSKDTGMTLKVRYSDSTAHNLILKSAFEKILREMGD TVTNSEREVNPGETHDTSTPKSFTLGTVLPSEKRNTTGDKLIRAGVPKSYGTRNDIA IIWPPNKPIVLSILSNHDKDDTLIADATKIVLETLKVTNK

This protease sequence is used for *insilico* approach by using various bioinformatics tools to study the structure, structure-function relationships and phylogeny of protease.

Conserved domains and motifs of alkaline protease from *Bacillus cereus* strain S8

The conserved domain database based protein domain annotation of alkaline protease from *Bacillus cereus* strain S8 (APBCS8) identified a similar type of domain present in Beta-lactamase 2 with E-value of 1.41e13 which confirms that APBCS8 belongs to transpeptidase super family through protein sequence annotation (Fig.1). Further, motif scan tool has identified two motifs in APBCS8 and the motifs are at amino acid residues located between 161-204 and 29-204. The motifs (Fig.2) are similar to that of Beta-lactamase family. These results confirm that APBCS8 belongs to Beta-lactamase family and transpeptidase super family.

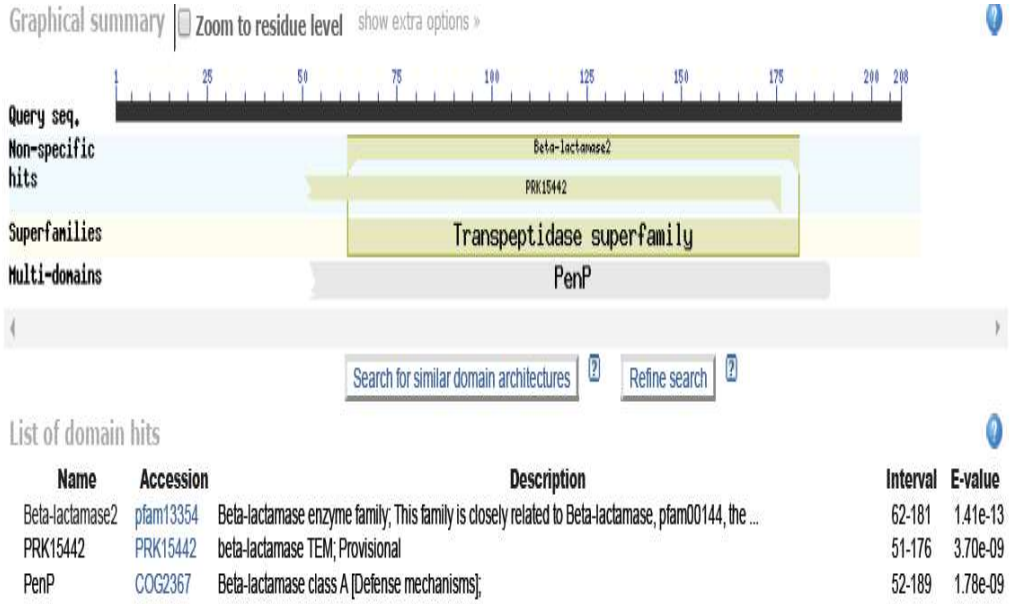


Fig. 1. Figure depicts maximum number of hits for Beta-lactamase 2 indicating that APBCS8 shared similar structural and functional properties of transpeptidase superfamily.

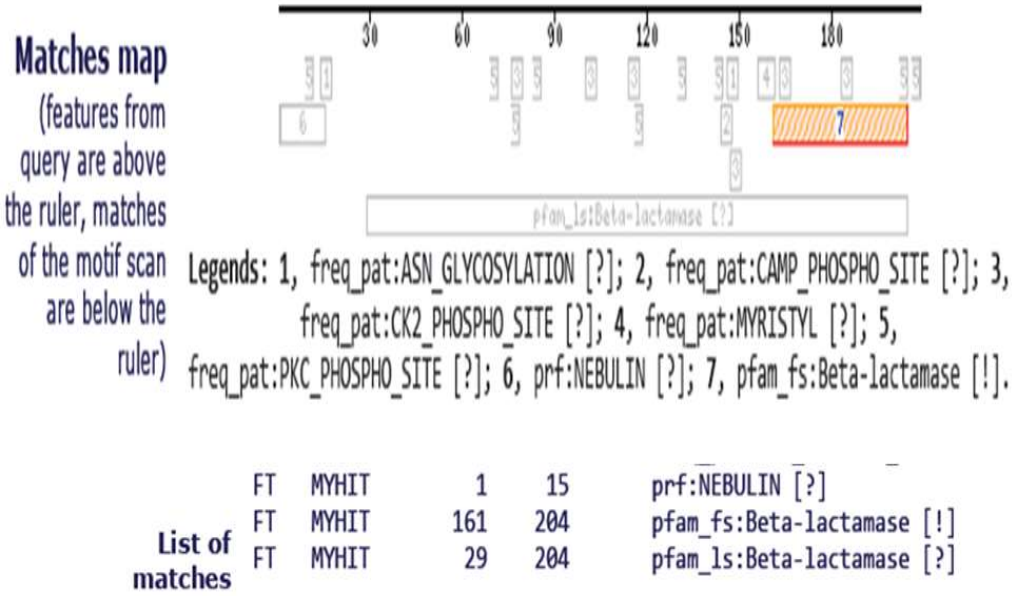


Fig. 2. Motif scan results for APBCS8.

Phylogenetic analysis

For the construction of phylogenetic tree of proteases, amino acid sequences of different types of proteases were obtained from protein database and p-BLAST. The phylogenetic tree (Fig.3) clearly indicates that serine alkaline protease of *Bacillus cereus* strain S8 (APBCS8) showed highest similarity with beta-lactamase 3 of *Bacillus cereus* ATCC 10876 with 100% nodal values. In the query of fourteen different proteases, it was observed that four proteases (mentioned as 3,2,4,5 in the phylogenetic tree) of different *Bacillus* species also belong to beta-lactamase family, and they showed close similarity as sister proteases. Serine β-lactamases are evolutionary related and belong to a superfamily transpeptidases, which also includes DD-peptidases. These proteins contain a Ser-x-x-Lys motif, where serine is the active site residue. Apart from this, it was observed that two proteases (11& 14) also appeared to be sister proteases which appeared as a sister clade. Proteases 12 and 13 of the phylogenetic tree were noticed to be a distinct lineage whereas proteases 9,8,7,10 were evolved as individual clades.

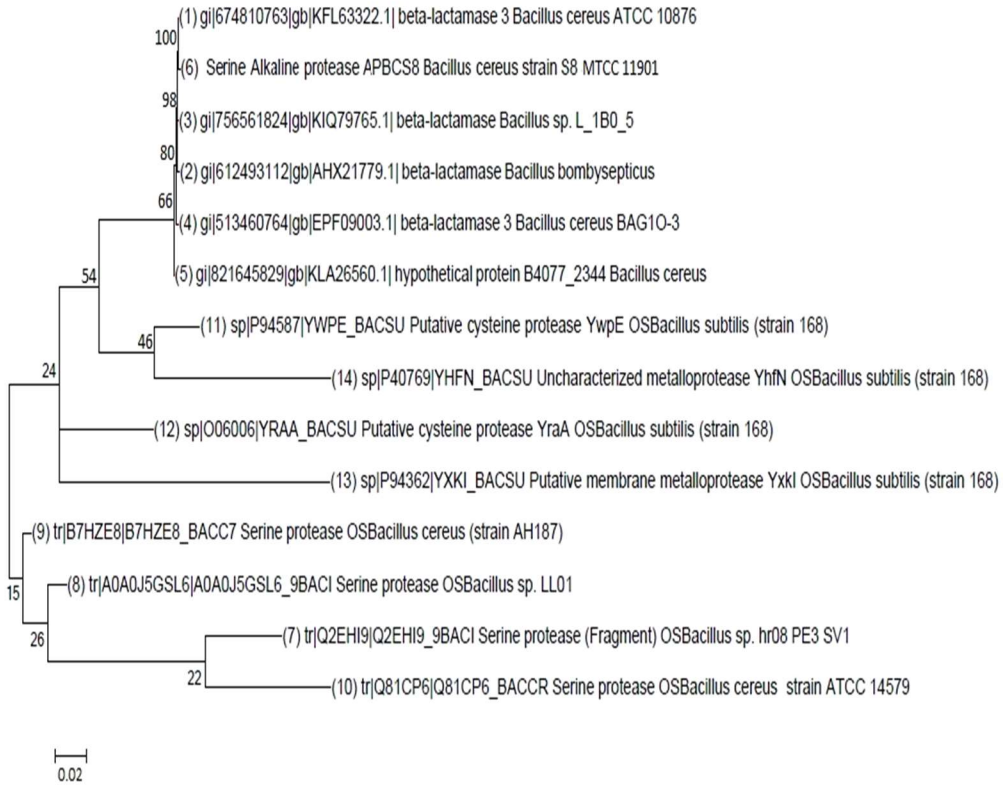


Fig. 3. Phylogenetic tree of proteases from different species

Modeling of proteins using LOMETS

Once a protein sequence has been determined, deducing its unique three dimensional (3D) native structure is a daunting task. Experimental methods to determine detailed protein structure, such as X ray diffraction studies and nuclear magnetic resonance (NMR) analyses, are highly labour intensive and the gap in the number between sequenced proteins and known structures increases. Thus, model building of proteins is of great importance. Knowing the structure of a protein sequence enables us to probe the function of the protein, understand substrate and ligand binding, devise intelligent mutagenesis and biochemical protein engineering experiments that improve specificity and stability, perform rational drug design, and design novel proteins.

The LOMETS threading has initially predicted secondary structure of proteins by PSI-PRED. The predicted secondary structure of the above mentioned identified protein by PSI-PRED was presented in Fig.4 (a). The server has predicted 10 best models after the initial evaluation, and the 3D model which was ranked first was generated. LOMETS has predicted 10 best models based on the confidence scores for alkaline protease S8 (Table 1), the 3D model of the protease was generated by using SP3 based on template 3v3sa (Table 2). The generated 3D structure was also uploaded in the PDB Sum database at http://www.biochem.ucl.ac.uk./bsm/ for further structural details (PDB assigned a code number m150 for Protease APBCS8).

Three-dimensional (3D) protein structure of the identified protein, alkaline protease from *Bacillus cereus* strain S8 (Fig.5) was generated using LOMETS, a threading based structure prediction method. The "confidence score" column indicates the confidence of the predicted template which is based on a scoring function that takes into account the z-score of the template, the confidence of the particular server and the sequence identity between the query and the template. Side chain contacts describe the predicted side chain contacts between pairs of residues and the confidence of the contact, defined as the relative frequency of contacts appearing in the top 20 templates. C- alpha atom contacts describe the predicted C- alpha atom contacts between pairs of residues, "short range C- alpha atom distances" describe the predicted C- alpha distances in short range, i.e i-j<7 and "Long range C- alpha atom distances" describe the predicted C- alpha distances in long range, i.e i-j>6. The modeled secondary structure of protease contains 3 beta sheets, 4 beta hairpins, 1 beta bulge, 9 strands, 10 helices, 9 helix helix interacs, 17 beta turns and 2 gamma turns (Fig. 4 (b))

```
Seq: FTNSHYKKSNKQTNQTNQVKQENKRNHAFAKLEKEYNAKTVAYHADDRAVGALLRQNSIE
SS : CCCCCCCCCHHHHHHHHHHHHHHCCCCEEEEEECCCCCEEEEECHHHHHHHHHHHHCCHH
Pos: 12345678901234567890123456789012345678901234567890123456789

Seq: ALDERITYTRKAFASTSKDTGMTLKVRYSDSTAHNLILKSAFEKILREMGDTVTNSEREV
SS : HCCCEEEECHHHHCCCCCCCCCEEEEEEECCHHHHHHHHHHHHHHHHHCCCCCCCCCCCC
Pos: 12345678901234567890123456789012345678901234567890123456789

Seq: NPGETHDTSTPKSFTLGTVLPSEKRNTTGDKLIRAGVPKSYGTRNDIAIIWPPNKPIVLS
SS : CCCCCCCCCCHHHHHHHCCCCHHHCCCCHHHHHHCCCCCCCCCCCEEEEECCCCCCEEEE
Pos: 12345678901234567890123456789012345678901234567890123456789

Seq: ILSNHDKDDTLIADATKIVLETLKVTNK
SS : EEECCCCCHHHHHHHHHHHHHHHHHCCC
Pos: 1234567890123456789012345678
```

Fig. 4 (a) Predicted secondary structure of Alkaline protease S8 by PSI-PRED.

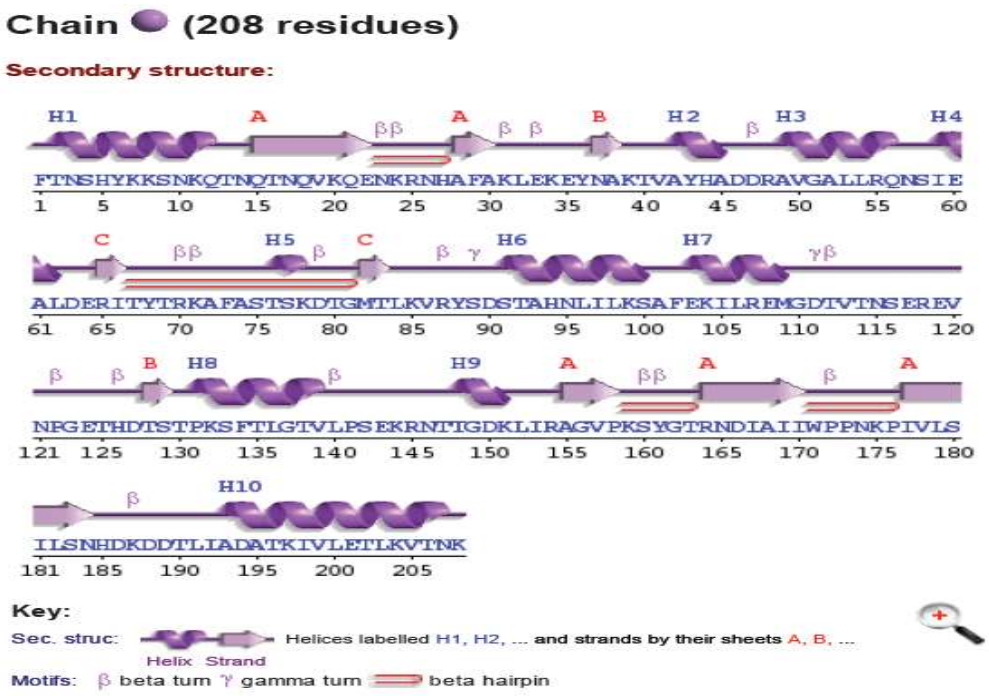


Fig. 4 (b) Secondary structure of Alkaline protease S8 generated by PDB sum.

Table: 1. Top 10 models selected by confidence score for Alkaline protease (*Bacillus cereus* strain S8).

Rank	Template	Align length	Coverage	Z score	Seq id	Confidence score	Program	Full length models by MODELLER
1	3v3sa	201	0.966	30.652	0.25	High	SP3	Model1
2	4dxbA	178	0.855	117.200	0.20	High	PRC	Model2
3	3v3sa	201	0.966	24.715	0.25	High	CdPPAS	Model3
4	3lezA	193	0.927	7.938	0.45	High	PROSPECT2	Model4
5	5hw3A	201	0.966	60.500	0.21	High	FFASO3	Model5
6	2gdn_A	193	0.927	79.931	0.38	High	FFAS-3D	Model6

7	4m3kA0	192	0.923	13.816	0.50	High	pGen THREADER	Model7
8	2qpnB	198	0.951	15.450	0.23	High	SPARKS-X	Model8
9	1ghp_A	189	0.908	20.916	0.34	High	HHSEARCH	Model9
10	4b88A	191	0.918	10.854	0.33	High	MUSTER	Model10

Table: 2. Models generated by SP3

Rank	Template	Align length	Coverage	Z score	Seq id	Confidence score
1	3v3sa	201	0.966	30.652	0.25	High
2	3ni9a	200	0.961	28.640	0.23	High
3	3niaa	199	0.956	28.395	0.24	High
4	3p09a	192	0.923	26.244	0.27	High
5	1bsg_	194	0.932	25.135	0.32	High
6	1iyoa	192	0.923	24.858	0.31	High
7	4ewfa	200	0.961	23.988	0.20	High
8	1ghpa	189	0.908	23.807	0.30	High
9	1w7fa	189	0.908	23.520	0.53	High
10	4mx4a	194	0.932	23.393	0.27	High

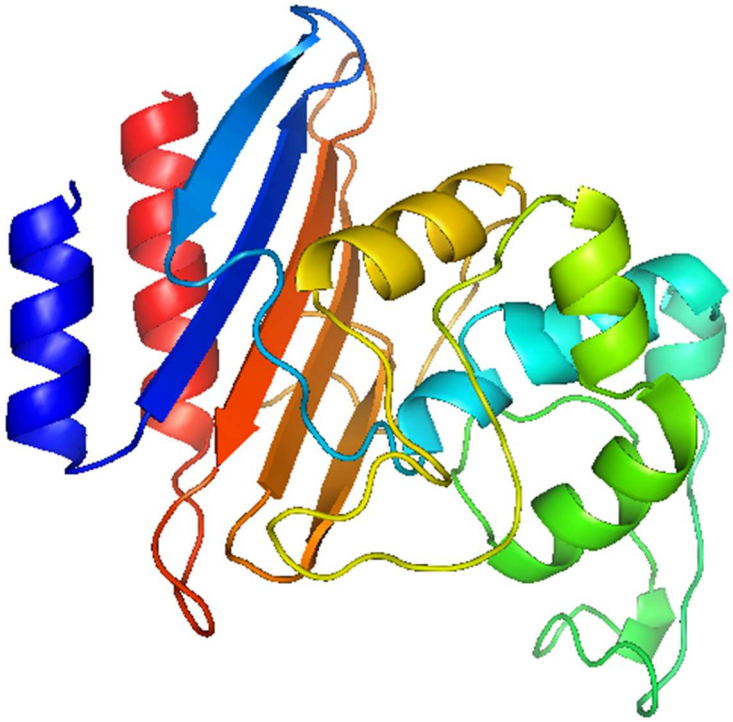


Fig. 5. Three dimensional structure of alkaline protease APBCS8.

Assessment and validation of the model

Alkaline protease from *Bacillus cereus* strain S8 (APBCS8) structure was further validated using PROCHECK server to assess possible conformation by Ramachandran plot. In Ramachandran plot (Fig. 6), the most allowed regions are indicated by red patches, while yellow areas show additional allowed regions and has an average allowed number of residues is 97.9% within the allowed region and 2.1% of the residues are in the disallowed region of the plot. Assessment and authentication results from Ramachandran plot analysis showed that alkaline protease from *Bacillus cereus* strain S8 (APBCS8) is of good quality.

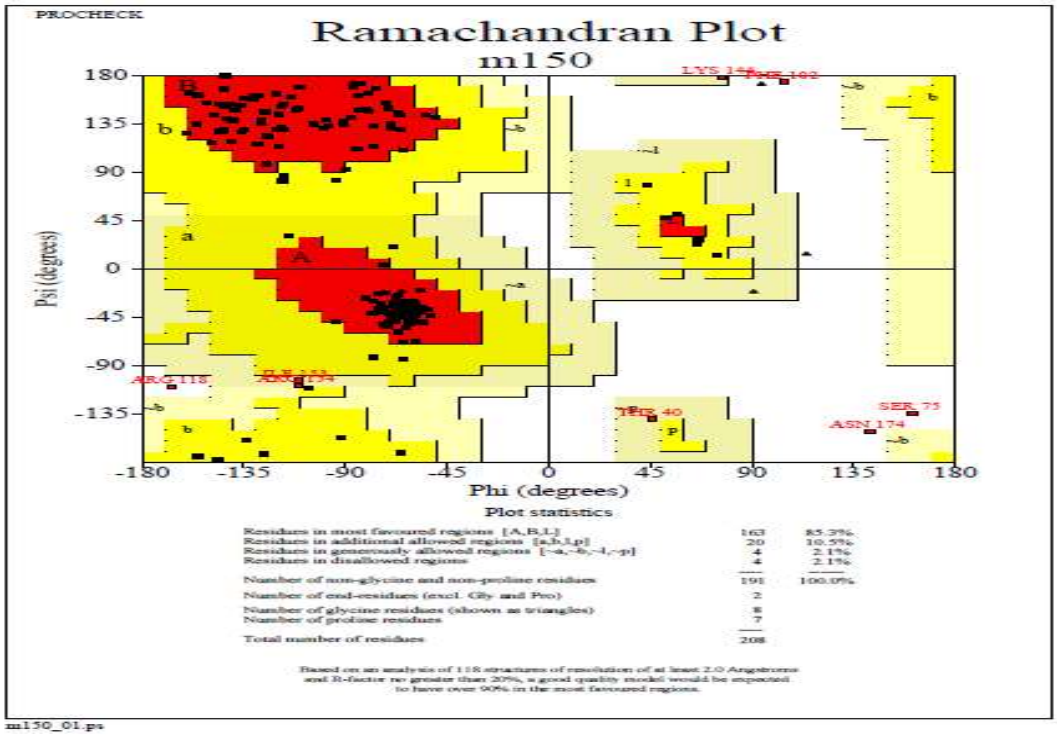


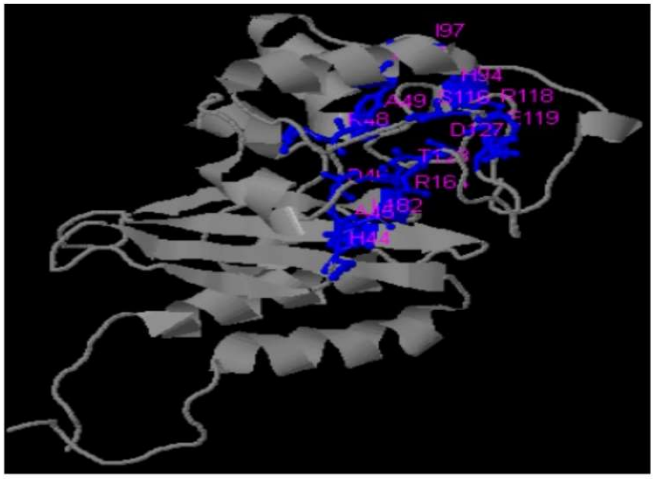
Fig. 6. Ramachandan plot of alkaline protease APBCS8.

According to the study carried out by Shaghayegh *et al*., (2015), structural modeling of a potential htrA-like serine protease from *Bacillus subtilis* DR8806 by MODELLER module of Discovery Studio software revealed that structure was a mixture of ß-strands and α-helix which was similar to other members of htrA family with two conserved domains including protease domain and PDZ like domain. Murugesan *et al*., (2016) carried work on three dimensional structure of metalloprotease from a phytopathogenic fungus, *Alternaria solani* by using DS MODELLER and proclaimed that it contains 9 α-helix and 2 ß-sheet, with identical topology.

Prediction of active site of APBCS8

A new reliable algorithm COACH (**CO**nsensus **A**pproa**CH**) was developed to predict ligand binding sites using the BioLiP database. COACH combines the results of five state of the art ligand binding sites prediction methods. Out of those, the result given by the ConCavity, one of the ligand binding site prediction method was found to be significant due to its highest C score value (0.28) (Fig 7) among the other prediction methods (Table 3). Highest C score value indicates the more reliability of the model. According to this method, it was predicted that active site (ligand binding site) was formed by the following residues: alanine (45), glutamic acid (119), isoleucine (97), threonine (164), phenylalanine (102), arginine (118), **histidine** (44, 94, 127), **aspartic acid** (46, 128), **serine** (116, 182) as shown in Figure 7. It was familiar that serine proteases mediate their mechanism of action by the involvement of catalytic triad contributed by the amino acids

histidine, aspartic acid and serine. Presence of these residues in the active site of alkaline protease from *Bacillus cereus* strain S8 strongly suggests that this protease is a serine protease, probably mediates its mechanism of action through catalytic triad. According to Chakraborty *et al*., (2011) shrimp alkaline phosphatase active site was detected by spatial conformity and electrostatic analysis by using CLASP. Jianyi *et al*., (2013) reported active site prediction and ligand interactions of phosphatidylinositol 4, 5-bisphosphate 3-kinase protein by a consensus approach (COACH).



C-Score	Predicted binding site residues by concavity
0.28	44,45,46,48,49,94,97,102,116,118,119,127,128,164,182

Fig. 7. The ConCavity ligand-binding sites prediction results. The confidence score (C score) and the predicted binding site residues are presented in the Table.

Table: 3. The output of top five predictions of other methods are briefly summarized below

Method	Rank	C-Score	Predicted binding site residues
COACH Results	1	0.18	84, 53,75,76,89,91,148,164
	2	0.10	166,168,170,179,181,192,193,197,200,201
	3	0.08	52,53,56,75,76,89,91,148,164
	4	0.08	148,164,189,192,193
	5	0.04	52,53,75,76,89,91,164
	6	0.03	75,187,189
	7	0.03	52,55,56,59,97,101,105,108
	8	0.01	53,89,148,164
	9	0.01	145,171,174
	10	0.01	6,7,8,9,10
TM-Site results			

	1	0.18	52,55,56,59,97,101,105,108
	2	0.14	148,164,189,192,193
	3	0.12	33,36,175
S-Site results			
	1	0.18	52,53,56,75,76,89,91,148,164
	2	0.19	149,152,153,156,157,158,164,166,168,170,179,181,189,192,193,196,197,200,201
	3	0.10	52,55,56,57,96,97,100,101,103,104,105,107,108,112
	4	0.10	75,76,152,185,186,187,188,189,190
	5	0.09	161,172,173,174,175,203
Co Factor Results			
	1	0.20	179,181,192,193,197,200,201
	2	0.14	181,193,196,197,200
	3	0.10	166,168,170,190,193,196
	4	0.06	52,55,56,59,97,101,105,108
	5	0.04	55,56,104,105,108
FINDSITE Results			
	1	0.15	50,77,80,89,91,120,161,162,163,164
	2	0.07	152,153,155,156,158,166,168,170,179,181,192,193,196,197,199,200,201
	3	0.06	151,152,155,161,162,164,166,189,192,193
	4	0.02	51,108,109,110,137,138,139
	5	0.01	157,171,173

Protein protein interactions

In the present study, a possible protein-protein interactome map for alkaline protease from *Bacillus cereus* strain S8 (belongs to beta-lactamse family) was constructed using STRING. A query of alkaline protease sequence was entered in to protein search box on start page of STRING website by selecting orthologs *Bacillus cereus* ATCC14579, which were considered as functional equivalent representatives and mapped on to the STRING network for alkaline protease. It was observed that alkaline protease (beta-lactamase represented by bla1) showed interactions with ten other different proteins (Fig. 8) associated with different functions in different pathways. Table 4 provides generalized information regarding the parameters involved in protein association of alkaline protease (beta-lactamase represented by bla1) whereas Table 5 provides detailed information regarding the function of each associated protein. As this protease belongs to beta-

lactamse family, most of the associated proteins involved in penicillin metabolic pathway, which is a generalized function of beta-lactamases. However it also shared associations with metalloprotease (ybeY), ATP-dependent clpP protease (clpP) and serine peptidylhydrolase (BC_2474). Apart from this, it also showed significant association with proteins involved in sporulation and hemolysis. Tables 6 and 7 provide information about its biological and molecular functions which was predicted by gene annotation, whereas Table 8 represents KEGG pathway. The alkaline protease from *Bacillus cereus* strain S8 with beta lactamase activity was supported by previous works. Protease with beta-lactamase activity from *Bacillus licheniformis* was reported previously by Aiyyappa *et al*., (1977). Asano *et al*., (1996) also reported an alkaline D-stereospecific endopeptidase with beta-lactamase activity from *Bacillus cereus*.

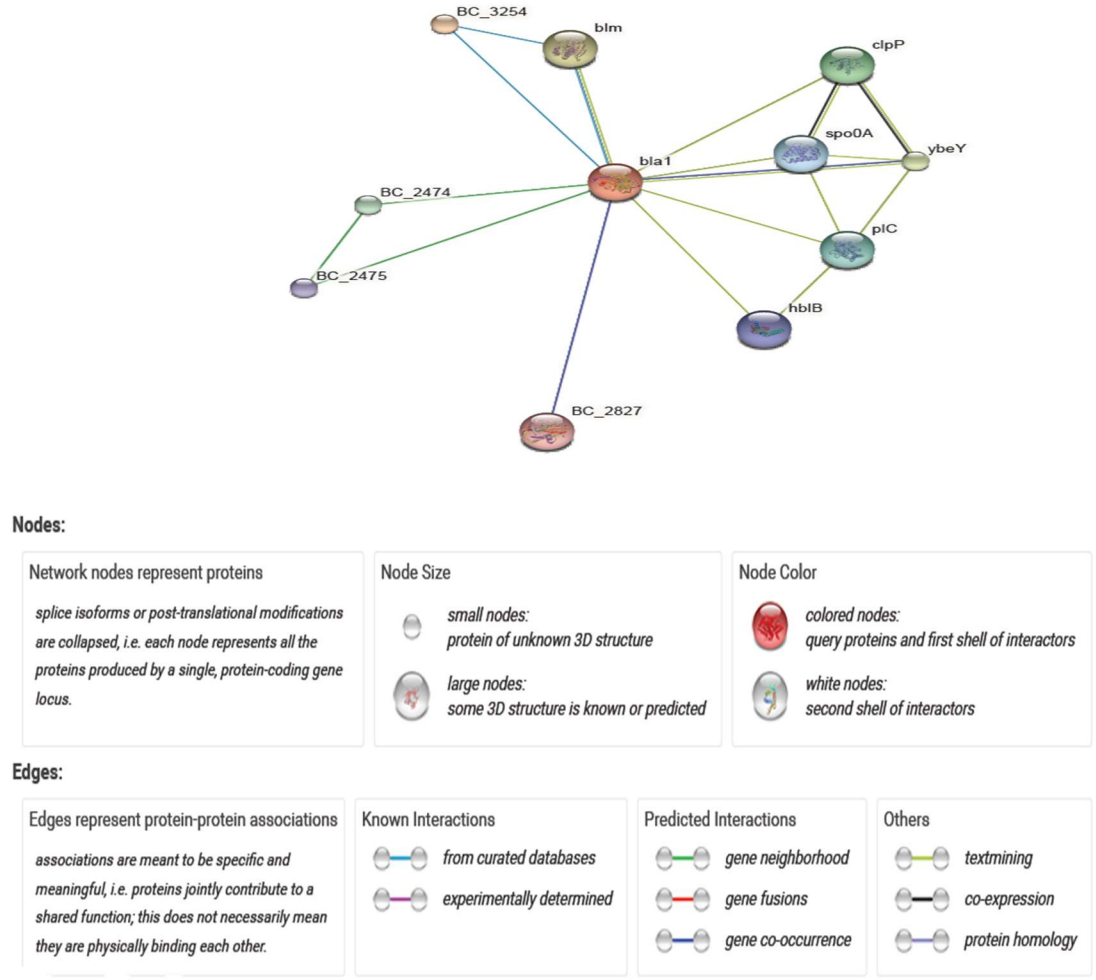


Fig. 8. Protein interactions of alkaline protease (Beta-lactamase).

Table: 4. Network Status general information

number of nodes	11
number of edges	18
average node degree	3.27
clustering coefficient	0.865
expected number of edges	10
PPI enrichment p-value	0.0179

Table: 5. Predicted Functional Partners

Protein	Function	Score
BC_3254	penicillin acylase II (763 aa)	0.900
blm	Beta-lactamase, type II; Can hydrolyze carbapenem compounds (264 aa)	0.806
ybeY	metalloprotease; Single strand-specific metallo-endoribonuclease involved in late-stage 70S ribosome quality control	0.584
clpP	ATP-dependent Clp protease proteolytic subunit; Cleaves peptides in various proteins in a process that requires ATP hydrolysis.	0.563
BC_2474	protein-serine peptidylhydrolase.	0.527
plC	phospholipase C; Required, with sphingomyelinase, to effect target cell lysis (hemolysis) (283 aa)	0.527
spo0A	stage 0 sporulation protein A; May play the central regulatory role in sporulation. It may be an element of the effector pathway	0.525
hblB	hemolysin BL binding component precursor (466 aa)	0.522
BC_2475	hypothetical protein (33 aa)	0.496
BC_2827	chitin binding protein (221 aa)	0.468

Table : 6. Biological process (GO)

Pathway ID	Pathway description	Count in gene set	False discovery rate
GO:0016999	antibiotic metabolic process	2	0.00157
GO:0017001	antibiotic catabolic process	2	0.00157
GO:0017144	drug metabolic process	2	0.00157
GO:0008150	biological_process	7	0.0176
GO:0046677	response to antibiotic	2	0.0176

Table : 7. Molecular Function

Pathway ID	Pathway description	Count in gene set	False discovery rate
GO:0008800	beta-lactamase activity	2	0.00187
GO:0016787	hydrolase activity	5	0.00187
GO:0008270	zinc ion binding	3	0.00531
GO:0016812	hydrolase activity, acting on carbon-nitrogen	2	0.00531
GO:0004175	endopeptidase activity	2	0.0165

Table : 8. KEGG Pathways

Pathway ID	Pathway description	Count in gene set	False discovery rate
00311	Penicillin and cephalosporin biosynthesis	3	7.3e-07
00312	beta-lactam resistance	2	0.0357

CONCLUSION

In this research, the 3D structure of alkaline protease from *Bacillus cereus* strain S8 was forecasted and evaluated using different bioinformatics tools. *In silico* analysis by LOMETS revealed the three dimensional structure of alkaline protease from *B. Cereus* strain S8. Conserved domains and motifs scan search revealed that this protease belongs to transpeptidase superfamily. Presence of catalytic triad (histidine (44, 94, 127), aspartic acid (46, 128), serine (116, 182) at active site of alkaline protease from *Bacillus cereus* strain S8 strongly suggests that this protease is a serine protease with potent industrial applications due to its unique characteristics. The enzyme's classification as a serine protease with a well-defined catalytic triad makes it a valuable

candidate for further research and development, holding the promise of addressing industrial and environmental challenges and contributing to the field of biotechnology.

Future scope

Exploring the ability of alkaline protease from B. Cereus strain S8 to degrade or detoxify organic pollutants or industrial waste can have environmental applications. Computational analysis can help predict its potential in bioremediation processes. Given the enzyme's unique characteristics and its classification as a serine protease, it may serve as a target for drug design and development. Computational methods can facilitate the identification of small molecule inhibitors that can be used in therapeutic applications. Further experimental studies, guided by in silico predictions, can be conducted to validate the enzyme's structure and its functional characteristics. This can include crystallography, site-directed mutagenesis, and kinetic studies. In silico analysis has laid the foundation for the exploration of alkaline protease from B. Cereus strain S8 in various industrial and biotechnological applications.

Acknowledgments

We, the authors are thankful to the Department of Biochemistry Andhra University, and Andhra Pradesh for providing the necessary amenities to carry out the present research work.

References

- **Van der Geer, J., Hanraads, J.A.J., Lupton, R.A., 2010. The art of writing a scientific article. J. Sci. Commun. 163, 51–59. https://doi.org/10.1016/j.Sc.2010.00372.**
- Gupta, R., Beg, Q. K., Khan, S., 2002a. An overview on fermentation, downstream processing and properties of microbial alkaline proteases. Appl. Microbiol. Biotechnol. 60, 381-395. http://link.springer.com/article/10.1007%2Fs00253-002-1142-1.
- Kirk, O., Borchert, T. V., Fuglsang, C. C., 2002. Industrial enzyme applications. Curr. Opin. Biotechnol.13, 345-351.http://www.sciencedirect.com/science/article/pii/S0958166902003282.
- Dias, D. R., Vilela, D. M., Silvestre, M. P. C., Schwan, R. F., 2008. Alkaline protease from Bacillus sp. isolated from coffee bean grown on cheese whey. World J. Microb. Biot. 24, 2027-2034. http://link.springer.com/article/10.1007%2Fs11274-008-9706-6.
- Jaouadi, B., Ellouz, C.S., Rhimi, M., Bejar, S., 2008. Biochemical & molecular characterization of detergent stable serine protease from *B. pumilus* CBS with high catalytic efficiency. Biochimie., 90: 1291-1305. doi: 10.1016/j.biochi.2008.03.004.

- Lopez, G., 2011. Firestar advances in the prediction of functionally important residues. Nucleic Acids Res., 39: 235-241. doi:10.1093/nar/gkr437.
- Roy, A., Zhang, Y., 2012. Recognizing protein-ligand binding sites by global structural alignment and local geometry refinement. Structure., 20: 987-997. doi: 10.1016/j.str.2012.03.009.
- An, J., 2005. Pocketome via comprehensive identification and classification of ligand binding envelopes. Mol. Cell. Proteomics., 4: 752-761. doi:10.1074/mcp.M400159-MCP200.
- Schmidt, T., 2011. Assessment of ligand-binding residue predictions in CASP9. Proteins., 79: 126-136. doi: 10.1002/prot.23174.
- **Lakshmi, B.K.M.,** Hemalatha, K.P.J., 2016. Production of Alkaline Protease from *Bacillus licheniformis* through Statistical Optimization of Growth Media by Response Surface Methodology. Fermentation Technology, 5(2): 2-7. DOI:10.4172/2167-7972.1000130.
- **Lakshmi B.K.M,** Ratna Sri P.V, Ambika Devi K and Hemalatha K.P.J, (2014). Screening, optimization of production and partial characterization of alkaline protease from haloalkaliphilic *Bacillus* sp. International Journal of Research in Engineering and Technology, 3:435-445.
- **Lakshmi, B.K.M**, Muni kumar, D., Hemalatha, K.P.J., 2018. Purification and characterization of alkaline protease with novel properties from *Bacillus cereus* strain S8. Journal of Genetic engineering and Biotechnology.
- Thompson, R.L., Mead, C.M., Edwards, R., 2004a. Effective bulk shears in super cell thunderstorm environments. Preprints, 22nd Conf. on Severe Local Storms, Hyannis, MA, Amer. Meteor. Soc. *https://www.spc.noaa.gov/publications/thompson/effshear.pdf.*
- Rzhetsky, A., Nei, M., 1992. A simple method for estimating and testing minimum evolution trees. Molecular Biology and Evolution 9:945-967. *https://pdfs.semanticscholar.org/9ebd/07d09b648b25b4581167493effda2031676a.pdf.*
- Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences (USA) 101:11030-11035. doi:10.1073/pnas.0404206101.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol., 24:1596-1599. doi:10.1093/molbev/msm092.
- Nei, M., Kumar, S., *2000.* Molecular evolution and phylogenetics. Oxford University Press, New York (333 pp). *https://global.oup.com/.../molecular-evolution-and-phylogenetics-9780195135855.*

- Sigrist, C.J.A., 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform. 3:265-274. https://www.ncbi.nlm.nih.gov/pubmed/12230035.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 4:406-425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.
- Huynen, M.A., Snel, B., Von Mering, C., Bork, P., 2003. Function prediction and protein networks. Curr. Opin. Cell Biol., 15: 191-198. https://www.ncbi.nlm.nih.gov/pubmed/12648675.
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., Hattori, M., 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res., 32: D278-D281. doi: 10.1093/nar/gkh063.
- Shaghayegh, F., Ahmad, A., Milad, L., 2015. Purification, biochemical characterization and structural modeling of a potential htrA-like serine protease from *Bacillus subtilis* DR8806. J. Mol. Cat. B: Enz., 115: 51-58. 10.1016/j.molcatb.2015.02.001.
- Murugesan, Ch., Raman, C., Se-Chul, Ch., Muthukrishnan, S., 2016. Isolation, characterization and molecular three dimensional structural predictions of metalloprotease from a phytopathogenic fungus, *Alternaria solani*. J. Biosci. Bioeng., 3: 1-9. doi: 10.1016/j.jbiosc.2015.12.021.
- Chakraborty, S., Minda, R., Salaye, L., Bhattacharjee, S.K., Rao, B.J., 2011. Active site detection by spatial conformity and electrostatic analysis-unravelling a proteolytic function in shrimp alkaline phosphatase. PLoS ONE 6(12):1-9. https://doi.org/10.1371/journal.pone.0028470.
- Jianyi, Y., Ambrish, R., Yang, Z., 2013. BioLiP: a semi-manually curated database for biologically relevant ligand–protein interactions. Nucleic Acids Research, 41: 1096-1103. doi: 10.1093/nar/gks966.
- Aiyappa, P.S., Traficante, L.J., Lampen, J.O 1977. Penicillinase-releasing protease of *Bacillus licheniformis*: purification and general properties. J. Bacteriol., 129(1):191-197. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC234914/.
- Asano, Y., Ito, H., Dairi, T., Kato, Y., 1996. An alkaline D-stereospecific endopeptidase with beta-lactamase activity from *Bacillus cereus*. J. Biol. Chem., 271(47):30256-30262. https://www.ncbi.nlm.nih.gov/pubmed/8939979.