ISSN (online): 2319-6599
ISSN (Print): 2319-6580

Sequence and Homology modeling analysis of Alkaline Protease from *Bacillus* sp

Md. Jahangir Alam

Dept. of Microbiology, Jessore University of Science and Technology, Country, Bangladesh
alamm5408@gmail.com

Abstract

Objectives: This study has been focused on the functional analysis and structure prediction of alkaline protease.
Methods: The study was performed on alkaline protease, sequence of which was taken from NCBI. The study includes functional sites analysis, physicochemical properties analysis, secondary structure and three dimensional structure predictions along with its validation.The Solvent Accessible Surface Area (SASA) of the protein was also measured. All these analysis and predictions were carried out using various computational tools.
Findings: These tools showed that the protein has 25 functional sites including myristoylation site, phosphorylation site, and glycosylation sites. It was found that the protein is stable, thermos table, hydrophobic/membranous and intracellular. 31.28% random coil structure confirms its flexibility while rest of the protein has its regular structure,confirming the stability of the protein. All of the validation results support that the predicted model of the protein was reliable.
Application: The study will be helpful for further research in signal transduction,intrinsic biological functions, and protein-protein/protein-ligand interaction and for designing new drug against it.
***Keywords*:** Bacillus, Alkaline Protease, Homology Modeling, SASA, Subtilisin, Myristoylation.

1. Introduction

Protease group is one of the most precious industrial enzymes responsible for the hydrolysis of peptide bonds. There are 62 clans (super family) and 268 families of protease enzyme, stored in MEROPS database [1]. Proteases are sub-divided into seven families based on the nature of the catalytic residues such as serine, cysteine, threonine, aspartic, glutamic, metalloprotease and asparagine peptide lyase [2]. Proteases are also classified into exopeptidase and endopeptidase based on the site of action. Exopeptidases are further grouped into amino-peptidase and carboxypeptidase. Endopeptidase includes serine, aspartic, cysteine, metallo, glutamic acid, and threonine proteases [3]. There are three forms of protease inclusive of acidic, neutral and alkaline proteases, classified on the basis of their activity and stability at pH. However alkaline proteases are the most usually used enzyme in industries for their surprisingly excessive activity and stability at alkaline pH [3] [4]. The genus *Bacillus* is an important source of commercial alkaline proteases and is probably the only genera being commercialized for alkaline protease production [5] [6]. Globally, proteases alone make contributions 60% marketplace proportion amongst all industrial enzymes, Bacillus being the major producer [7]. Microbial alkaline proteases are extensively utilized in special industries which include detergent, peptide synthesis, fabric, leather-based, silver healing, dairy, baking, drinks, and pharmaceutical industries [3] [8-11].

The enzyme protease is involved in wide varieties of function in human body. It digests dietary protein to allow absorption of amino acids in the small bowel. Other functions include blood coagulation immune function maturation of prohormones bone formation and programmed cell death. Proteases also offer a valuable target in many therapeutic settings including Alzheimer's cancer and viral infection [12]. Most of the function of the protein is mediated by the catalytic triad. Catalytic triad is a combination of three coordinated amino acids (aspartate, histidine and serine) that is a major component of the active site of the protein.

It performs catalysis through a nucleophilic attack to covalently bind with the substrate protein breaking into two half of the product [13]. The aim of this study was to analyze the primary and secondary structure of alkaline protease of *Bacillussp.* Along with its functional sites, to model 3D structure of alkaline protease by using homology modeling tools and to validate its model by using various bioinformatics tools.

2. Material and Methods

2.1. Sequence retrieval

The amino acid sequence of alkaline protease was retrieved from NCBI (https://www.ncbi.nlm.nih.gov/protein). The protein sequence with the Accession number "CAE51830.1" was selected for the present study.

2.2. Functional site prediction

CD-search [14], Scan Prosite [15], Inter Pro Scan [16] and Motif Scan [17] tool was used to determine the functional regions such as domain, motif, and active sites, present in alkaline protease. In this study, the FASTA sequence of the selected proteins was uploaded and the results obtained were analyzed.

2.3. Physicochemical properties analysis

Prot Param [18] is an online tool which is available on the ExPASy server. It computes various physicochemical properties that can be deduced from a protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY). In this study, the FASTA sequence of the selected proteins was uploaded and the results attained were analyzed.

2.4. Trans-membrane segment analysis

The TMHMM [19] server is an online tool which can clearly predict trans-membrane helices and capable of discriminating between soluble and membrane proteins with both sensitivity and specificity. In this study, the FASTA sequence of the query protein was pasted onto the input section of the server and rest of the parameters were set at default.

2.5. Secondary structure prediction

In this current study, to attain quantitative values for the number of alpha-helices, beta sheets and coils present within the amino acid stretch of the protein, the Self Optimized Prediction Method (SOPMA) tool [20] was used. The FASTA sequence of the query protein was pasted onto the input section of the server.

2.6. Disulfidebond prediction

Three online programs (DISULFIND [21], CYSPRED [22] and DiANNA [23])were used to identify SS-bonding states of cysteines and location of disulfide bridges of proteins. In this study, the query protein FASTA sequence was entered and then computed. The results attained were then recorded and analyzed.

2.7. Structure prediction

Homology modeling of the query protein (CAE51830.1)was performed using Modeler 9.19 software [24], Swiss-Model [25] and M4T server [26]. The Basic Local Alignment Search Tool (BLAST) [27] finds regions of local similarity between sequences. In this study, BLASTp was used against protein data bank. 1c9m for Swiss model was selected as template sequence as it was the nearest homologous sequence (82.46% sequence identity) with lowest E.value. 1gci was selected as template automatically in M4T server which has also a good similarity with the query protein. Modeler is command based software. Four PDB files were selected as template for Modeler.

2.8. Validation

After modeling, the quality and validation of the model were evaluated by several structure assessment methods, containing Z-Score by using QMEAN [28], Rampage Ramachandran plot analysis [29] and ERRAT [30]. In order to evaluate the reliability of the predicted structure of alkaline protease, the root mean square deviation (RMSD) was measured by superimposing it on the template structure using Pymol software [31].

2.9. Solvent Accessible Surface Area (SASA) Analysis

VADAR, an online program, was used to measure solvent exposed area of the protein. Residue wise SASA was calculated through GETAREA online server

3. Results and Discussion

3.1. Functional site analysis

Several tools/servers were used for functional site prediction such as CD-Search, InterProScan, Motif Scan, and ScanProsite. CD search result [concise result] showsthat its identity with the subtilisin-like family belonging to super family subtilases. Limited proteolysis of most large protein precursors is executed in vivo with the use of subtilisin-like pro-protein convertases. Many essential biological functions including peptide hormone synthesis, viral protein processing, and receptor maturation involve proteolytic processing by using these enzymes [32]. CD search result shows two domains, first of which is Peptidases_S8_Subtilisin_subset. This group is consisting of many different subtilisins such as Pro-TK-subtilisin, subtilisin Carlsberg, serine protease Pb92 subtilisin, and BPN subtilisins. In pro-TK-subtilisin, Ca^{2+} binding is responsible for stability,folding and activity of protein [33] [34]. Subtilisin Carlsberg possesses the highest commercial importance as a proteolytic additive for detergents [35]. Second is Peptidase inhibitor I9, known to function as molecular chaperones, assisting in the folding of the mature peptidase [36].

Table 1. The motifs of Alkaline Protease by Motif Scan

Motif information	No. of sites	Amino acid residues
N-glycosylation site.	5	160-163, 179-182,226-229, 317-320,358-361
Casein kinase II phosphorylation site.	1	57-60
N-myristoylation site.	13	130-135, 164-169, 173-178, 186-191, 200-205, 209-214, 221-226, 230-235, 257-262, 303-308, 310-315, 318-323, 365-370
Serine proteases, subtilase family, aspartic acid active site.	1	133-143
Serine proteases, subtilase family, histidine active site.	1	167-177
Serine proteases, subtilasefamily,serine active site.	1	318-328
Curlin associated repeat	1	270-292
Peptidase inhibitor I9	1	29-106
Subtilase family	1	111-365

There were nine different motifs observed in the MotifScan result (Table1). N-linked protein glycosylation affects protein function, stability, and protein complex formation in three domain of life [37]. Phosphorylation is a significant process which affects the functions and activities of proteins and modulates the cell behavior through controlling its intrinsic biological activity, cellular localization, and interaction with other proteins [38]. N-myristoylation group play a vital role in signal transduction by enhancing dynamic protein-protein and protein-membrane interactions [39]. Both phosphorylation and myristoylation are post-translational protein modification, adding phosphate group and myristic acid in protein respectively. Both processes contribute to structural stability of protein [40] [41].

The thirteen N-myristoylation sites, five N-glycosylation site and one phosphorylation site in the query sequence support that my protein is structurally stable for playing role in signal transduction, protein-protein interaction, and intrinsic biological functions.Along with Motif Scan result, Scan Prosite and Inter pro Scan results show that there are three active sites(aspartic, histidine, and serine) in the query protein. These amino acid residues are the three constituents of catalytic triad in serine protease and thus involved in the process of blood clotting, protein digestion, cell signaling, inflammation, and protein processing [42].

3.2. Physico-chemical properties analysis

The physicochemical properties of alkaline protease were analyzed using Protparam. The result showed that (Table 2) this protein had 374 amino acids with molecular weight of 38286.52 Daltons and pI of 4.60

Table 2. The physicochemical properties of Alkaline Protease

Parameter	Value	Explanation
Isoelectric point (pI)	4.60	The protein is accepted as acidic.
Total number of negatively charged residues (Asp + Glu) and total number of positively charged residues (Arg + Lys).	35 and 17 respectively	The total numbers of negatively charged residues are higher than the total number of positively charged residues. This protein has few intracellular portions.
The instability index (II)	24.75	This classifies the protein as stable.
Aliphatic index	88.48	This indicates that this protein is thermos table.
Grand average of hydropath city (GRAVY)	0.040	A positive GRAVY score reveals that the protein is hydrophobic/ membranous.

The most abundant amino acid was found as alanine (51 residues, 13.6%), whereas the lowest was cysteine (0 residues, 0%). The total number of negatively charged residues (Asp + Glu, 35) was found higher than the total number of positively charged residues (Arg + Lys, 17). The protein containing higher numbers of negatively charged residues than positively charged residues indicate the protein as intracellular [43] [44] [45]. Intracellular proteins have the lower number of cysteine residues, but also higher numbers ofaliphatic and charged amino acid residues [46]. This data is in agreement with our finding that the highest number of the amino acid residue was alanine, while the lowest one was cysteine. The aliphatic index, which is defined as the relative volume of a protein occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine), of proteins of thermophilic bacteria is significantly higher than that of ordinary proteins [47]. My protein is stable as its instability index (II) is 24.75, far less than 40. It has been reported that all the unstable proteins have an II > 40, whereas all the stable proteins, with the only exception of RNaseA, have an II < 40 [48]. It has been reported that proteins that have an in vivo half-life of less than 5h shows instability index of more than 40, whereas those that have an in vivo half-life of more than 16 h [49]. Hydrophobicity score (arbitrary unit) below 0 are more likely globular (hydrophilic protein), while scores above 0 are more likely membranous (hydrophobic) [50]

3.3. Trans-membrane segment analysis

There was no trans-membrane helix in this protein sequence. But the expected number of amino acids in trans-membrane helices is 2.2827. If this number is larger than 18 and number of transmembrane helices is one or more, it is very likely to be a transmembrane protein (or have a signal peptide) [51]. So this protein is not a trans-membrane protein.

3.4. Secondary structureprediction

The secondary structure of the protein was predicted using SOPMA server (Table 2). It was observed that alpha helix was predominant (34.76%) followed by the random coil (31.28%), and extended strand (21.12%) and beta-turn (12.83%). This finding could be related to the enzymatic function of the protein. Linus Pauling and Robert Corey proposed two periodic structures called the α helix (alpha helix) and the β pleated sheet (beta pleated sheet) as common types of secondary protein structure [52]. Random coils are important in flexibility and conformational changes of the protein [53]. Since 31.28% of this protein has random coiled structure, the protein may have enzymatic function. Rest of the protein has αhelices (alpha helix) and β sheets which are theprotein's regular secondary structure, confirming the stability of the protein. The stability of this protein indicates that it can retain its structural conformation or its activity when subjected to physical or chemical treatment. The SOPMA result is shown in Table 3.

ISSN (online): 2319-6599
ISSN (Print): 2319-6580

Table 3. Secondary structure of Alkaline Protease using SOPMA

Parameters	Number of amino acids	Amino acids (%)
Alpha helix	130	34.76%
Extended strand	79	21.12%
Beta-turn	48	12.83%
Random coil	117	31.28%

3.5. Disulfidebond prediction

As there was no cysteine residue in the query sequence, there was no sulfur atom. Therefore, there was no disulfide bond in the sequence.

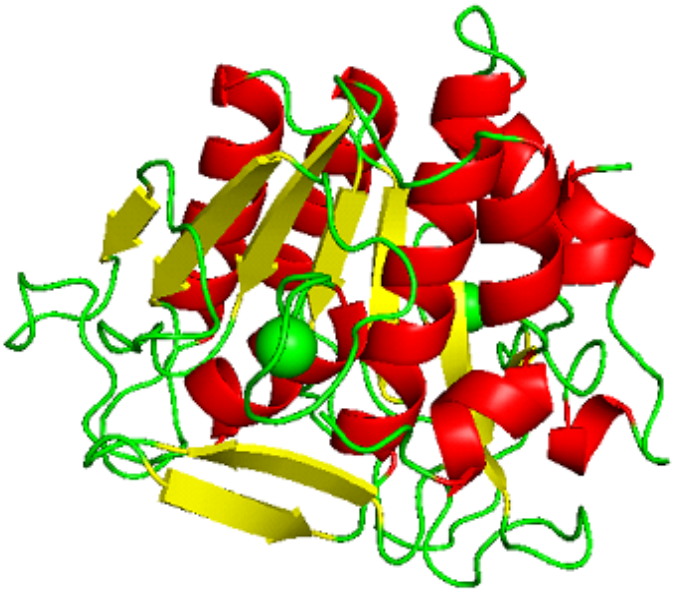
3.6. BLAST Search and Alignment between Target and Template Sequence

The query sequence was run in BLAST against Protein Data Bank (PDB). BLAST result showed that the query protein is highly similar with "1c9m" with lowest E-value (2E-157). 1c9m was chose as template for predicting protein structure of the query sequence. The template protein has 269 amino acids while the querysequence has 374 amino acids (Figure 1).

Figure 1. Query-Template Sequence Alignment Using Clustal Omega

```
Temlate     ------------------------------------------------------------  0
Query       MGKIVAGTALIISVAFSSSIAQAAEEAKEKYLIGFKEQEVMSQFVDQIDGDEYSISSQAE  60

Temlate     ---------------------------------------------AQSVPWGISRVQAPA  15
Query       DVEIDLLHEFDFIPVLSVELDPEDVDALELDPAIAYIEEDAEVTTMQTVPWGINRVQAPI  120
                                                          *:*****.*****

Temlate     AHNRGLTGSGVKVAVLDTGISTHPDLNIRGGASFVPGEPSTQDGNGHGTHVAGTIAALDN  75
Query       AQSRGFTGTGVRVAVLDTGISNHADLRIRGGASFVPGEPNISDGNGHGTHVAGTIAALNN  180
            *:.**:**:**:*********.* **.***********. .*****************:*

Temlate     SIGVLGVAPSAELYAVKVLGASGSGAISSIAQGLEWAGNNGMHVANLSLGSPSPSATLEQ  135
Query       SIGVLGVAPNVDLYGVKVLGASGSGSISGIAQGLQWAANNGMHIANMSLGSSAGSATMEQ  240
            *********.:**.**********:**.*****:**.*****:**:**** : ***:**

Temlate     AVNSATSRGVLVVAASGNSGAGSISYPARYANAMAVGATDQNNNRASFSQYGAGLDIVAP  195
Query       AVNQATASGVLVVAASGNSGAGNVGFPARYANAMAVGATDQNNNRASFSQYGAGLDIVAP  300
            ***.**: **************.:.:**********************************

Temlate     GVNVQSTYPGSTYASLNGTSMATPHVAGAAALVKQKNPSWSNVQIRNHLKNTATSLGSTN  255
Query       GVGVQSTVPGNGYASFNGTSMATPHVAGVAALVKQKNPSWSNVQIRNHLKNTATNLGNTT  360
            **.**** **. ***:************.************************.**.*.

Temlate     LYGSGLVNAEAATR      269
Query       QFGSGLVNAEAATR      374
             :************
```

Figure 2. Predicted model generated by Swiss-model server



Template sequence was aligned with the query protein using clustal omega. The alignment result showed that the 82% amino acids of both sequence (query and template) are identical. Conservative and semi-conservative sequence was 5.94% and 7.43% respectively. This identity and conservation confirms similarity that may indicates functional, structural and evolutionary relationships between two biological sequences [54].

3.7. Construction of homology models of target protein and its validation

Among five models generated in modeler v9.19, the model having lowest dope score was selected as final model. This model and another two model generated by Swiss-model server and M4T server were validated. Quality and reliability of structure were checked by several structure assessment methods including Rampage, ERRAT, and Q Mean-Score by various parameters and Ramachandran plot analysis. From this analysis, Swiss-model result is observed to be most reliable result (Figure 2).

Table 4. Summary of successfully produced models by using Modeller v9.19

Filename	Molpdf	DOPE score	GA341 score
M1.pdb	41866.69922	-22382.30078	1.00000
M2.pdb	52466.47656	-20275.80664	1.00000
M3.pdb	25919.28711	-27431.52539	1.00000
M4.pdb	44525.96875	-22779.67188	1.00000
M5.pdb	50181.96875	-22338.47461	1.00000

The stereo chemical quality of the modeled protein was evaluated by the use of RAMPAGE. Ramachandran plot evaluation of the model predicted from Swiss model shown in Table 4,5 that 1.9% of amino acid residues in the allowed region and 98.1% in the favored region, indicating that the models were of most reliable and best quality.

ERRAT is a protein structure verification algorithm that analyzes statistics of non-bonded interactions between different atom types based on characteristic atomic interaction [55]. ERRAT result of the model generated from Swiss model showed that the overall quality factor was found as 96.429 which are more reliable and very satisfactory than those from M4T and Modeller.

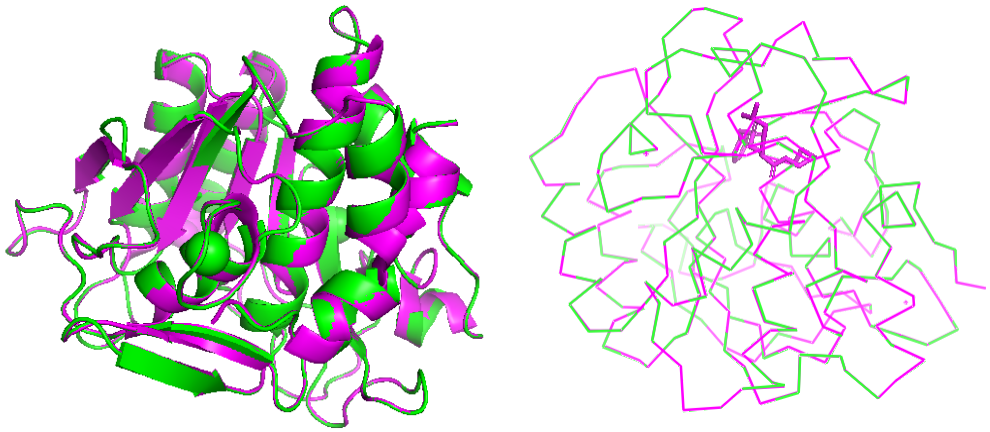
Table 5. Comparative values of RAMPAGE, ERRAT, Q Mean scores and RMSD between the template and models generated by Modeller, Swiss-Model and M4T server

	RAMPAGE	ERRAT	QMEAN Score	RMSD
Modeller9.19	Favored region =89.8% Allowed region =4.6% Outlier region =5.6%	40.667%	-8.19	3.0073
Swiss-Model	Favored region =98.1% Allowed region =1.9% Outlier region =0.0%	96.4286%	0.41	0.089
M4T	Favored region =97.8% Allowed region =2.2% Outlier region =0.0%	95.402%	-0.45	0.16

For QMEAN scores, the estimated model reliability is between 0-1 with higher values indicating more reliable candidates. The QMEAN score of the model generated from Swiss model showed was 0.40 which is very reliable and indicates the absolute quality of themodel. This result is far better than that of M4T and Modeller.Among three tools, the Root Mean Square Deviation (RMSD) result from Swiss-model was lowest and it was the best. The Root Mean Square Deviation (RMSD) between the template and predicted model of query protein was calculated by superimposing it on the template using Pymol software and the RMSD value was 0.08976 (Figure 3). It suggests that the predicted model has strong homology with the template. The low RMSD between the template and predicted model of query protein confirms strong homology between them [56].

ISSN (online): 2319-6599
ISSN (Print): 2319-6580

Figure 3. a) Superposition of template (magenta) with the predicted model of alkaline protease (green) and b) Superposition of backbone traces. The picture has been taken from Pymol software



a) b)

3.8. Solvent Accessible Surface Area (SASA) Analysis

VADAR result shown in Table 6 that total volume of the protein is 31796.1Angs**3 and exposed area is 9414.9 Angs**2, and rest of the protein is buried/hydrophobic. Residue wise SASA was calculated through GETAREA online server. Charged and polar residues showed higher solvent accessible surface area than other hydrophobic residues. Although a separate weren't much exposed. Both charged and polar residues play a vital role in binding free energy. These residues indicate that they may act as binding site cavities. Hydrophobic residues, constituting the interior surface of the protein, correlate well with hydrogen exchange rates and thermo-stability at mutagenesis condition. It also measure in the prediction of protein–protein interaction hot spots detecting sites for post-translational modification and predicting the impact of point mutations on protein stability and function [57].

Table 6. Solvent Accessible Surface Area Analysis from VADAR and Get area

(a) Results from VADAR			(b) Results from Getarea	
Parameter	Surface Area		Polar and charged Amino Acids with residue number	Surface Area (Angs**2)
Total volume	31796.1Angs**3		ASN-355	117.00
Total ASA	9414.9 Angs**2		GLN-240	117.45
Exposed nonpolar ASA	5084.6 Angs**2		SER-339	108.00
Exposed Polar ASA	3466.2 Angs**2		THE-359	100.24
Exposed charged ASA	864.1 Angs**2		ARG-374	185.58
			LYS-336	66.31
			ASP- 280	26.06
			GLU-239	63.15

4. Conclusion

The functional sites and physicochemical analysis of the protein were evaluated using various computational tools. This analysis will support to understand the roles and biological effects of the protein. As there were thirteen N-myristoylation sites, five N-glycosylation sites and one phosphorylation site observed in the query protein, the protein can be used for the research in signal transduction, protein-protein interaction, and intrinsic biological functions. Model predicted has a good reliability. So, this model will help to design ligand and to understand structure-function relationship of the protein. The Solvent Accessible Surface Area (SASA) was also measured. It allows understanding about ligand binding sites. As it is only a predicted model, further experiment (*in vivo and in vitro*) should be done to confirm its structure.

5. References

1. N.D. Rawlings, A.J. Barrett, P.D. Thomas, X. Huang, A. Bateman, R.D. Finn. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*. 2018; 46(1), 624–632.
2. K. Oda. New families of carboxyl peptidases: Serine-carboxyl peptidases and glutamic peptidases. *Journal of Biochemistry*. 2012; 151(1), 13–25.
3. M.B. Rao, A.M. Tanksale, M.S. Ghatge. Molecular and Biotechnological Aspects of Microbial Proteases. *Microbiology and Molecular Biology Reviews*. 1998; 62(3), 597–635.
4. S. Guleria, A. Walia, A. Chauhan, C.K. Shirkot. Purification and characterization of detergent stable alkaline protease from Bacillus amyloliquefaciens SP1 isolated from apple rhizosphere. *Journal of Basic Microbiology*. 2016; 56(2), 138–152.
5. H. Varela, M. Ferrari, L. Belobrajdic, A. Vazquez, L. Loperena. Skin unhairing proteases of Bacillus subtilis: Production and partial characterization. *Biotechnology letters*. 1997; 19(8), 755-758.
6. G. Kocher, S. Mishra, Immobilization of Bacillus circulans MTCC 7906 for enhanced production of alkaline protease under batch and packed bed fermentation conditions. *The Internet Journal of Microbiology*. 2010; 7(2), 1-6.
7. W.C.A. Do Nascimento, M.L. Leal Martins. Production and properties of an extracellular protease from thermophilic Bacillus sp. *Brazilian Journal of Microbiology*. 2004; 35(1–2), 91–96.
8. A. Anwar, M. Saleemuddin. Alkaline proteases: A review. Bioresource Technology. 1998; 64, 175–183.
9. P. Ellaiah, B. Srinivasulu, K. Adinarayana. A review on microbial alkaline proteases. *Journal of Scientific & Industrial Research*. 2012; 61(9), 690–704.
10. K. Saeki, K. Ozaki, T. Kobayashi, S. Ito. Detergent alkaline proteases: enzymatic properties, genes, and crystal structures. *Journal of Bioscience Bioengineering*. 2007; 103(6), 501–508.
11. R. Singh, M. Kumar, A. Mittal, P.K. Mehta. Microbial enzymes: industrial progress in 21st century," *3 Biotechnology*. 2016; 6(2), 1–15.
12. J.M. Milner, A. Patel, A.D. Rowan. Emerging roles of serine proteinases in tissue turnover in arthritis. *Arthritis Rheum.* 2008; 58(12), 3644–3656.
13. L. Polgar. The catalytic triad of serine peptidases. *Cellular and Molecular Life Sciences*. 2005; 62(19–20), 2161–2172.
14. A. Marchler-bauer. CDD : a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*. 2011; 39(3), 225–229.
15. E. De Castro. ScanProsite : detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research*. 2018; 34(3), 362–365.
16. R.D. Finn. InterPro in 2017 — beyond protein family and domain annotations. *Nucleic Acids Research*. 2017; 45, 190–199,
17. B.T. Naughton, E. Fratkin, S. Batzoglou, D.L. Brutlag. A graph-based motif detection algorithm models complex nucleotide dependencies in transcription factor binding sites. *Nucleic Acids Research*. 2006; 34(20), 5730–5739, 2018.
18. E. Gasteiger. Protein Identification and Analysis Tools on the ExPASy Server BT - The Proteomics Protocols Handbook," J. M. Walker, Ed. Totowa, NJ: Humana Press. 2005; 571–607.
19. A. Krogh, È. Larsson, G. Von Heijne, E.L.L. Sonnhammer. Predicting tranmembrane protein topology with a Hidden Markov model : application to complete Genomes. *Journalof Molecular Biology*. 2001; 305(3), 567-80.
20. C. Geourjon, G. Deléage. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Computer Applications in the Biosciences*. 1995.
21. A. Ceroni, A. Passerini, A. Vullo, P. Frasconi. DISULFIND : a disulfide bonding state and cysteine connectivity prediction server. *Nucleic Acids Research*. 2006; 34, 177–181.
22. P. Fariselli, P. Riccobelli, R. Casadio. Role of evolutionary information in predicting the disulfide-bonding state of cysteine in proteins. *Proteins*. 1999; 36(3), 340-6.

23. P. Clote. DIANNA : a web server for disulfide connectivity prediction. *Nucleic Acids Research*. 2011; 33, 230–232.
24. A. Sali, T. Blundell, A. Sali, T.L. Blundell. Comparative modeling by satisfaction of spatial restraints. *Journalof Molecular Biology*. 1994; 234, 779-815.
25. M. Biasini. SWISS-MODEL: modeling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, 2014; 42, 252–258.
26. N. Fernandez-fuentes, C.J. Madrid-aliste, B.K. Rai, J. Eduardo. M4T: a comparative protein structure modeling server. *Nucleic Acids Research*. 2007; 35, 363–368.
27. S.F. Altschup, W. Gish, T. Pennsylvania, U. Park. Basic local alignment search tool. Department of Computer Science. *Journalof Molecular Biology*. 1990; 403–410.
28. P. Benkert, M. Biasini, T. Schwede. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*. 2011; 27(3), 343–350.
29. S.C. Lovell, I.W. Davis, W.B. Arendall, P.I.W. de Bakker, J.M. Word, J.S. Richardson, D.C. Richardson. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins Structure*. 2002; 50, 437–450.
30. D.S. Obukhov. Errat 2: Verification of protein structure. University of California, Los Angeles. 2002; 1-132.
31. DeLano, W. L. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on Protein Crystallography*. 2002; 40, 82–92.
32. F. Bergeron, R. Leduc, R. Day. Subtilase-like pro-protein convertases: From molecular specificity to therapeutic applications. *Journal of Molecular Endocrinology*. 2000; 24(1), 1–22.
33. S. ichi Tanaka, Y. Takeuchi, H. Matsumura, Y. Koga, K. Takano, S. Kanaya. Crystal structure of Tk-subtilisin folded without propeptide: Requirement of propeptide for acceleration of folding. *FEBS Letter*. 2008; 582(28), 3875–3878.
34. S. ichi Tanaka, H. Matsumura, Y. Koga, K. Takano, and S. Kanaya. Identification of the Interactions Critical for Propeptide-Catalyzed Folding of Tk-Subtilisin. *Journal of Molecular Endocrinology*. 2009; 394(2), 306–319.
35. P.J. Dor. Detection of antibodies to proteases used in laundry detergents by the radioallergosorbent test. *The Journal of Allergy and Clinical Immunology*. 1986; 78(5), 877–886.
36. Y. Li, Z. Hu, F. Jordan, M. Inouye. Functional analysis of the propeptide of subtilisin E as an intramolecular chaperone for protein folding. Refolding and inhibitory abilities of propeptide mutants. *Journal of Biology Chemistry*. 1995; 270(42), 25127–25132.
37. H. Nothaft, X. Liu, D.J. Mcnally, C.M. Szymanski. Functional Glycomics. *Methods in Molecular Biology*. 2010; 600(1), 227–243.
38. P. Cohen. The regulation of protein functions by multisite phosphorylation – a 25 year update. *Trends Biochemistry Science*. 2001; 25, 596–601.
39. N. Burnaevskiy. Proteolytic elimination of N-myristoyl modifications by the Shigella virulence factor IpaJ. *Nature.* 2013; 496, 1-106.
40. H.B. Olsen, N.C. Kaarsholm. Structural Effects of Protein Lipidation as Revealed by LysB29-myristoyl, des (B30) Insulin. *Biochemistry*. 2000; 39, 11893–11900.
41. J. Zheng, D.R. Knighton, S.S. Taylor, NH. Xuong, J.M. Sowadski, L.F. Ten Eyck. Crystal structures of the myristylated catalytic subunit of cAMP-dependent protein kinase reveal open and closed conformations. *Protein Science*. 1993; 2(10), 1559–1573.
42. Ö.D. Ekici, M. Paetzel, R.E. Dalbey. Unconventional serine proteases: Variations on the catalytic Ser/His/Asp triad configuration. *Protein Science*. 2008; 17(12), 2023–2037.
43. J. Masin. Negatively charged residues of the segment linking the enzyme and cytolysin moieties restrict the membrane-permeabilizing capacity of adenylate cyclase toxin. *Scientific Reports*. 2016.
44. S. Verma, V. Bednar, A. Blount, B.G. Hogue. Identification of Functionally Important Negatively Charged Residues in the Carboxy End of Mouse Hepatitis Coronavirus A59 Nucleocapsid Protein. *Journal of Virology*. 2006; 80(9), 4344–4355.
45. C. Rutz, W. Rosenthal, R. Schülein. A single negatively charged residue affects the orientation of a membrane protein in the inner membrane of Escherichia coli only when it is located adjacent to a transmembrane domain. *Journal of Biology Chemistry*. 1999; 274(47), 33757–33763.

46. H. Nakashima, K. Nishikawa. Discrimination of Intracellular and extracellular proteins using amino acid composition and residue-pair frequencies. *Journal of Molecular Biology*. 1994; 238(1), 54–61.
47. A. Ikai, Thermo stability and Aliphatic Index of Globular Proteins. *Journal of Biochemistry*. 1980; 88(6), 1895-1898.
48. K. Guruprasad, B.V.B. Reddy, M.W. Pandit. 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 Engineering, Design and Selection*. 1990; 4(2), 155–161.
49. S. Rogers, R. Wells, M. Rechsteiner. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science*. 1986; 234(4774), 364-368.
50. S. Magdeldin. Murine colon proteome and characterization of the protein pathways. *Bio Data Min*. 2012; 5(1), 1-11.
51. A. Krogh, B. Larsson, G. Von Heijne, E.L.L. Sonnhammer. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journalof Molecular Biology*. 2001; 305(3), 567–580.
52. J. Berg, J. Tymoczko, L. Stryer. Secondary Structure: Polypeptide Chains Can Fold Into Regular Structures Such as the Alpha Helix, the Beta Sheet, and Turns and Loops. Biochemistry. 2002, 1-1050.
53. N.C. Fitzkee, G.D. Rose. Reassessing random-coil statistics in unfolded proteins. Proc. Natl. Acad. Sci. U. S. A. 2004; 101(34), 12497–12502.
54. N. Behera, M.S. Jeevitesh, J. Jose, K. Kant, A. Dey, J. Mazher. Higher accuracy protein multiple sequence alignments by genetic algorithm. *Procedia Computer Science*. 2017; 108, 1135–1144.
55. C. Colovos, T. Yeates. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science*. 1993; 1511–1519.
56. V.N. Maiorov, G.M. Crippen. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *Journalof Molecular Biology*. 1994; 235(2), 625–634.
57. K.P. Tan, R. Varadarajan, M.S. Madhusudhan. DEPTH : a web server to compute depth and predict small-molecule binding cavities in proteins. *Nucleic Acids Research*. 2018; 39, 242–248.

The Publication fee is defrayed by Indian Society for Education and Environment (www.iseeadyar.org)
Cite this article as:
Md. Jahangir Alam. Sequence and Homology modeling analysis of Alkaline Protease from *Bacillus* sp. *Indian Journal of Bioinformatics and Biotechnology.* Vol 6 (6), June 2018.