International Journal of Basic and Applied Biology
Print ISSN: 2394-5820, Online ISSN: 2349-2539, Volume 2, Number 5; April-June, 2015, pp. 310-314
© Krishi Sanskriti Publications
http://www.krishisanskriti.org/ijbab.html

In-silico Comparative analysis of Papain Family Cysteine Protease using Computational Tools and Servers

Neetu Jabalia[1*], Hina Bansal[2], P.C. Mishra[3] and Nidhee Chaudhary[4]

[1,2,4]Amity University Uttar Pradesh, Noida
[3]Guru Nanak Dev University, Amritsar, Punjab
E-mail: [1]njabalia@amity.edu, [2]hbansal@amity.edu,
[3]pcm.biotech@gndu.ac.in, [4]nchaudhary@amity.edu

Abstract—*Enzymes have greatly contributed to environmentally adapted clean and green technology due to their biodegradable nature and therefore have replaced harsh chemicals to a great extent. Proteases are a unique class of enzymes as they possess both degradative and synthetic properties. Their applications in industry and therapeutics have grown rapidly in the last two decades. Proteases comprise a vast group of enzymes with applications in various industries viz; pulp and paper, textiles, detergent, leathering, baking and bioremediation processes. They account for 65% of total world wide enzyme sale. Multiple application of these enzymes stimulated interest to discover them with novel properties and considerable advancement of basic research into these enzymes. A broad understanding of the active site of the enzyme and of the mechanism of its inactivation is essential for delineating its structure-function relationship. In this investigation, seven papain family cysteine protease from different organisms were retrieved from NCBI. All the seven protease sequences were analyzed and characterized in silico using various computational Their physico-chemical properties, hydropathicity and secondary structure have been identified. Phylogenetic analysis was also performed using MEGA6. The study might be an initiation in understanding the underlying structural and functional aspects of these proteins, their common characteristics and other features for academic and industrial purposes.*

1. INTRODUCTION

Proteases (also termed as proteolytic enzymes or proteinases) refer to a group of enzymes whose catalytic function is to hydrolyze peptide bonds of proteins. These enzymes are widely distributed in all plants, animals and microorganisms. Proteases account for approximately 2% of the human genome and 1 to 5% of genomes of infectious organisms [1, 2]. The necessity of proteases has made them industrially important enzyme. They account for 65% of total worldwide enzyme sale. Many proteases have been isolated from latex, fruits and seeds. They have importance in both commercial and physiological fields. They have major applications in industrial process such as laundry, silk, pharmaceutical, food and degradation of gelatin on X-ray films [3]. They are also used in bioremediation process. They are widely used in detergents, leather, waste management and silver recovery [4].

Proteases are ubiquitous in nature. A variety of microorganisms such as bacteria, fungi and actinomycetes are known to yield these enzymes [5]. It would be beneficial to utilize a fungal protease as fungal expression systems are accomplished of producing huge quantities of enzymes [6]. Proteases are formed by various species of fungi such as *Aspergillus* [7], *Mucor* [8], *Fusarium* [9], *Cephalosporium* [10] *and Rhizopus* [11]. Filamentous fungi, such as *Aspergillus* are used for large scale production of bulk industrial enzymes. [12]. *Actinomycetes*, a protease producing bacteria used in industrial purpose, pharmaceutical and cytotoxic agent [13]. *Bacillus sp.* was found to be predominant and rich source of alkaline proteases and have been reported to produce extracellular alkaline protease [14]. *B. substilis* produce large amount of protease and used for industrial purposes [15]. Among several proteases, bacterial proteases are more significant compared to with animal and fungal protease [16]. They are present in *Bacillus subtilis*, *B.amyloliquefaciens, Pseudomonas sp., Lysobacter enzymogene, E. coli* etc. [17]. *B. subtilis* yield alkaline protease that can be utilize in textiles, leather and food industries [18]. Plants are also the rich source of protease enzyme [19]. The objective of this study is to characterize all the sequences of papain family cysteine protease from *Legionella pneumophila*, *Dictyocaulus viviparus*, *Pandoravirus inopinatum*, *Bacillus cereus, Gregarina niphandrodes, Ancylostoma duodenale* and *Capsaspora owczarzaki* ATCC 30864 by using *in-silico* techniques which will be valuable to understand the structural features and will raise the prospects of its academic or commercial use.

2. MATERIAL AND METHODS

2.1 Sequence analysis

Papain family cysteine protease sequences were retrieved from the NCBI (National Center for Biotechnology Information) by using the keywords 'Papain family cysteine protease' for *Legionella pneumophila*, *Dictyocaulus viviparus*, *Pandoravirus inopinatum*, *Bacillus cereus, Gregarina niphandrodes, Ancylostoma duodenale* and *Capsaspora owczarzaki* ATCC 30864 (Table 1).

2.2 Physico-chemical characterization

The amino acid compositions (Fig. 1) of papain family cysteine protease sequences were computed using the Expasy's Protparam tool. The physico-chemical characterization, isoelectric point (Ip), molecular weight (M.W.), total number of positive and negative residues, extinction coefficient (E.C.), instability index (I.I.) [21], aliphatic index (A.I) [22] and grand averge hydropathicity (GRAVY) [23] were computed using the Expasy ProtParam [24] prediction server (Fig. 2).

Table 1: Papain family cysteine protease sequences retrieved from NCBI.

S. No.	Accession No.	Sequence description	Organism name	Sequence length
1	GAN27912	papain family cysteine protease	Legionella pneumophila	364 AA
2	KJH4665	papain family cysteine protease	Dictyocaulus viviparus	356 AA
3	YP_009120189	papain family cysteine protease	Pandoravirus inopinatum	370 AA
4	AJI16751	papain family cysteine protease	Bacillus cereus	268 AA
5	XP_011129855	papain family cysteine protease	Gregarina niphandrodes	372 AA
6	KIH68116	papain family cysteine protease	Ancylostoma duodenale	214 AA
7	KJE88537	papain family cysteine protease	Capsaspora owczarzaki	589 AA

2.3 Secondary structure prediction

The analysis of the secondary structure of protein sequences were based only on knowledge of their primary structure. The secondary structure feature of the papain family cysteine protease sequences from *Legionella pneumophila*, *Dictyocaulus viviparus*, *Pandoravirus inopinatum*, *Bacillus cereus, Gregarina niphandrodes, Ancylostoma duodenale* and *Capsaspora owczarzaki* ATCC 30864 were identified by GORIV [25]. It is used to describe secondary structure freatures such as sequence length, alpha helix, beta turn and random coil etc. (Fig. 3).

2.4 Phylogenetic Analysis

Phylogentic tree (Fig. 4) was constructed using neighbor-joining method [26] using Molecular Evolutionary Genetics Analysis (MEGA) software version 6 [27].

3. RESULTS AND DISCUSSION

The amino acid sequence determines the fundamental properties of the enzymes. A set of conserved amino acid residues located in vicinity that provide clues to the functions is termed as motif. Motifs were predicted using Motif Finder. It has been predicted that all the sequences have similar motif among all the protease but at different positions (Table 2). Protease sequences were classified as soluble protein by SOSUI server except KJH46615 and KJE88537.1, which were classified as membrane proteins. SOSUI server identified one transmembrane region in KJH46615 and KJE88537.1.

The amino acid composition of papain family cysteine protease is represented in Fig. 1**.** The values of isoelectric point (pI) of the proteases from *Legionella pneumophila*, *Dictyocaulus viviparus*, *Pandoravirus inopinatum*, *Bacillus cereus, Gregarina niphandrodes, Ancylostoma duodenale* and *Capsaspora owczarzaki* ATCC 30864 were in the range 4.96 to 9.18 indicating that all are acidic in nature except YP_009120189.1 and KIH68116.1 because their pI value was more than 7. The computed isolelectric point will be useful for developing buffer system when these enzymes are to be purified in solution by isoelectric focusing method [28]. The Expasy's Protparam was used to determine the extinction coefficient of the proteases. Extinction coefficient of proteases at 280nm ranged from 31525 to 126500 M^{-1} cm^{-1} with respect to the concentration of Cys, Trp and Tyr. The extinction coefficient was high in KJE88537.1 indicating the presence of high concentration of aromatic amino acids. The computed protease concentrations and extinction coefficients will be used in the quantitative study of protein-protein and protein-ligand interactions in solution.

The aliphatic index (AI), defined as the relative volume of a protein occupied by aliphatic side chains, is regarded as a positive factor for the increase of thermal stability of globular proteins. Aliphatic index for the protease sequences ranged from 59.72 – 91.65. The very high aliphatic index of these

sequences indicated that the cysteine protease will be stable over a wide temperature range [29]. The Grand Average hydropathy (GRAVY) value for a peptide or protease is calculated as the sum of hydropathy values of all the amino acids, divided by the total number of residues present in the cysteine protease sequence [30]. GRAVY indices for the sequences ranged from -0.5 to -0.046. This low range of value indicated better interaction with water (Fig. 2).

Secondary structural features were predicted by GORIV (Fig. 3). The results revealed that random coils dominated among secondary structure elements. The conformational entropy associated with random coils significantly contributes to stabilization and protein folding. Proline, which has a high content in the proteases, has special property of creating kinks in polypeptide chains and disrupting ordered secondary structure and might have contributed to the high content of random coil structure.

Table 2 : Motifs and hydropathicity predicted from the sequence.

Accession No.	Average hydropathicity	Motif position
GAN27912	-0.238462	290-347
KJH4665	-0.305899	91-345
YP_009120189	-0.045946	16-212
AJI16751	-0.414552	45-251
XP_011129855	-0.360753	66-174, 250-345
KIH68116	-0.500467	73-210
KJE88537	-0.097453	53-281, 351-580

A phylogenetic tree was constructed using the protein sequences of Legionella pneumophila, Dictyocaulus viviparus, Pandoravirus inopinatum, Bacillus cereus, Gregarina niphandrodes, Ancylostoma duodenale and Capsaspora owczarzaki ATCC 30864 (Fig. 4) through Neighbour joining methods using MEGA6, to clarify the phylogenetic relationship among all the protease sequence and infer their evolutionary history.

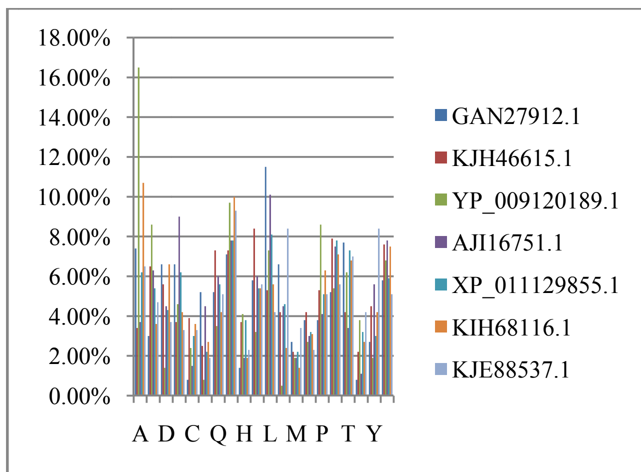

Fig. 1 : Amino acid composition of the protease sequences.

The tree has been evaluated by bootstrapping method. The analysis revealed that all the papain family cysteine protease sequences had originated from a common ancestor and during the course of evolution diverged further into sub groups. Papain family cysteine protease from B. cereus, D. viviparous and G. niphandrodes belongs to a one sub group while P. inopinatum, C. owczarzaki, A. duodenale and L. pneumophilia belong to the another sub group.

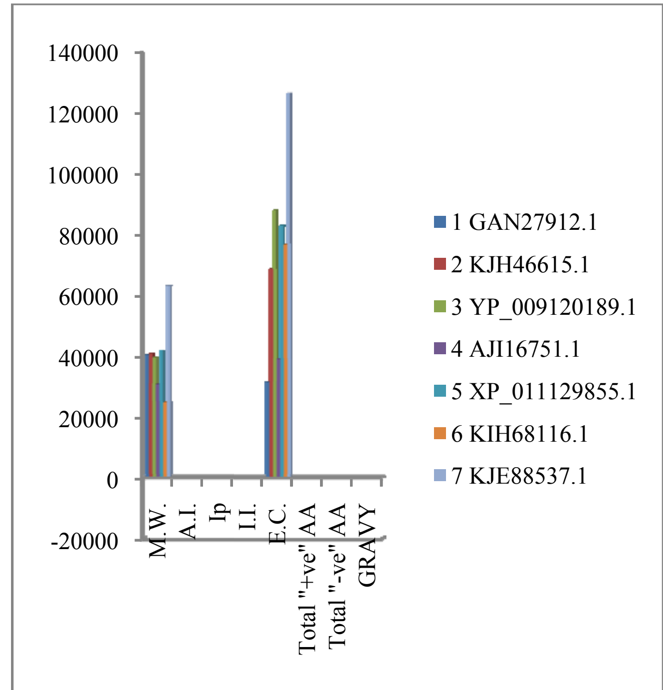

Fig. 2: Parameters computed using ExPASY's Protparam tool.

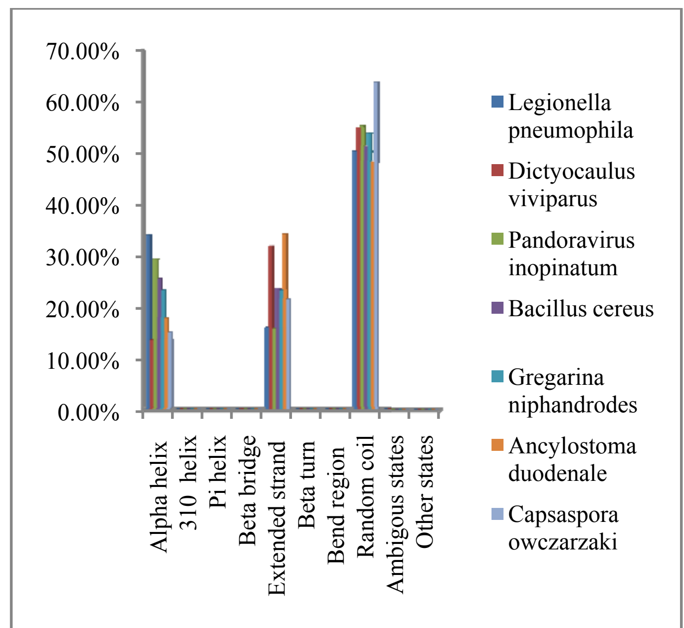

Fig. 3: Percentage of amino acid sequences forming secondary structure in GORIV prediction.

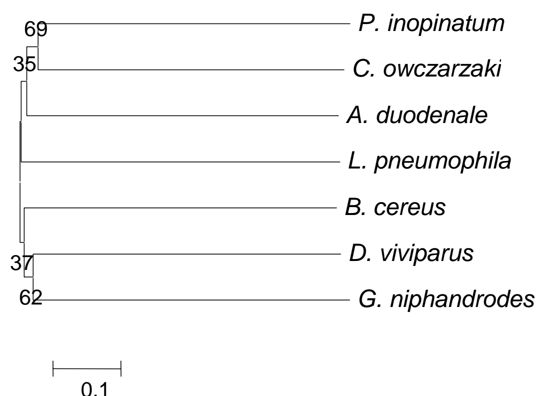

Fig. 4: Results of phylogenetic analysis using MEGA 6.

4. CONCLUSION

In our study, we have characterized amino acid sequences of multiple papain family cysteine proteases present in Legionella pneumophila, Dictyocaulus viviparus, Pandoravirus inopinatum, Bacillus cereus, Gregarina niphandrodes, Ancylostoma duodenale and Capsaspora owczarzaki ATCC 30864. Total seven sequences of cysteine protease have been analyzed to acquire an understanding about their functional properties, physico-chemical properties and phylogenetic analysis by using in silico techniques. By performing motif search it has been observed that papin family cysteine protease domain is present but at varied position. SOUSI predicts that all are the soluble protein except KJH46615 and KJE88537.1. Primary structure analyses revealed that the proteases were hydrophilic and are expected to be stable over wide range of temperature. Secondary structure analysis established that in most of the sequences, random coils were the dominating secondary structure elements followed by alpha helix, extended strand and beta turns. The construction of phylogeny tree was done by neighbor joining algorithm. The evolutionary analysis shows that all the protease sequences share common ancestor and they are evolutionary related. This study will provide insight about the physicochemical properties and function of cysteine protease which will further aid in formulating their uses in academics and industries.

5. ACKNOWLEDGMENT

We are grateful to our Director of Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida for his constant support and encouragement during this study.

REFERENCES

[1] Puente, X. S., Sanchez, L. M., Overall, C. M. and Lopez, O. C., "Human and mouse proteases: A comparative genomic approach", *Nauture Reviews Genetics*, 4, July 2003, pp. 544–548.

[2] Jabalia, N., Mishra, P. C. and Chaudhary N., "Applications, Challenges and Future Prospects of Proteases: An Overview", *Journal of Agroecology and Natural Resource Management*, 2014, pp. 179-183.

[3] Ishikawa, Ishimi, K., Sugiura, M., Sowa, A. and Fujiwara, N., "Kinetics and mechanism of enzymatic hydrolysis of gelatin layers of X-ray film and release of silver particles", *.J Ferm. Bioeng*. 76, 1993, 300-305.

[4] Sumantha, Sandhya, C., Szakacs, G., Soccol, C. R. and Pandey, A., " Production and partial purification of a neutral metalloprotease by fungal mixed substrate fermentation", *Food Technology and Biotechnology*, 2005, pp. 313-319.

[5] Madan, M., Dhillon, S. and Singh, R., "Production of alkaline protease by a UV mutant of *Bacillus polymyxa*", *Indian J. Microbiol*., 42, 2002, pp. 155-159.

[6] Anandan, D., Marmer, W. and Dudley, R., L., "Isolation, characterization and optimization of culture parameters for production of an alkaline protease isolated from *Aspergillus tamarii"*, *J.Ind. Microbiol. Biotechnol.*, 34, 2007, pp. 339-47.

[7] Chakraborty, Srinivasan, M. S. and Raghwan, S.K., "Production of acid proteases by a new *Aspergillus niger* during solid substrate fermentation", *J . Microbiool, Biotechnol*, 10, 1995, pp. 17-30.

[8] Thakur, M. S., Karant, N. G. and Nand, K., "Production of fungal rennet by *Mucor Miehi* using solid state fermentation", *Appl. Microbiol Biotechnol*, 32, 1990, pp. 409-413.

[9] Khan, M. R., Blain, J. A and Patterson, J. D. E., "Intracellular protease of *Mucor Pusillus* Lindt. Pak", *J. Biochem*, 1, 1981, pp. 1-8.

[10] Tuschiya, K., Arai, T., . Seki, K. And Kimeua, T., "Purification and some properties of alkaline proteinases from *Cephalosporium* sp. KM 388", *J. Argic, Biol. Chem*. 51, 1987, pp. 2959-2965.

[11] Ikasari, L. and Mitchell, D .A., "Protease production by *Rhizopus oligosposus* in solid state fermentation", *Microbiol Biotechnol*, 10, 1994, 320-324.

[12] Bergquiest, P. L., Te'O Jr. V. S., Gibbs, M. D., Cziferszky, A. C. E., Azevedo, M. O. and Nevalainen, K. M. H., Proudction of recombinant bleaching enzymes from thermophilic microorganisms in fungal hosts, *Appl Biochem Biotechnol*, 98-100, 2002, 165-176.

[13] Balachandran, Duraipandiyan, V. and Ignacimuthu, S., "Purification and characterization of protease enzyme from actinomycetes and its cytotoxic effect on cancer cell line (A549)", Asian Pacific Journal of Tropical Biomedicine, 2012, pp. 392-400.

[14] Joo, H. S. and Chang, C. S., "Production of an oxidant and SDS-stable alkaline protease from an alkaophilic *Bacillus clausii* I-52 by submerged fermentation: Feasibility as a laundry detergent additive", *Enzyme microbial*, 2006, pp. 176-183.

[15] Nisha, N. S. and Divakaran, J., " Optimization of alkaline protease production from *Bacillus substilis* NS isolated from sea water", *African Journal of Biotechnology*, 2014, pp. 1707-1713.

[16] Tunga, Shrivastava, B. and Banerjee, R., " Purification and characterization of a protease from solid state cultures of *Aspergillus parasiticus*", *Process Biochemistry*, 2003, pp. 1553–1558.

[17] Ustariz, F. J., Laca, A.,. Garcia, L. A and Diaz, M., "Fermentation of individual proteins for protease production by *Serratia marcescens", Biochem. Engr. J.*, 2004, pp. 147-153.

[18] Vanitha, N., Rajan, S. and Murugesan, A. G., "Optimization and production of alkaline protease enzyme from *Bacillus substilis* 168 isolated from food industry waste", *International journal of current microbiology and applied sciences*, 2014, pp. 36-44.

[19] Sharmila, J. S., Jeyanthi, R. L., Das, M. P. and Saduzzaman, Md., "Isolation and partial purification of Protease from plant

leaves*", Journal of Chemical and Pharmaceutical Research*, 2012, pp. 3808-3812.

[20] Hirokawa, T. "SOSUI: classification and secondary structure prediction system for membrane proteins", *Bioinformatics*,*14*, 1998, 378-379.

[21] Gill, S., C. and Von, Hippel, P. H., "Calculation of protein extinction coefficients from amino acid sequence data", *Anal. Biochem.182*, 1989, 319-326.

[22] Guruprasad, K., "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*. *4*, 1990, 155-161.

[23] Ikai, A.J., "Thermostability and aliphatic index of globular proteins", *J. Biochem*. *88*, 1980, pp. 1895-1898.

[24] Kyte, J. and Doolittle, R. F., "A simple method for displaying the hydropathic character of a protein", *J. Mol. Biol*. *157*, 1982, pp. 105-132.

[25] Gasteiger, E. and Walker John, M., "The Proteomics Protocols Handbook", Humana Press, 2005, pp. 571-607.

[26] Garneir, J., Gibrat, J. F. and Robson, B., "GOR secondary structure prediction method version IV", *Methods Enzymol. Doolittle Ed.* 266, 1996, pp. 540-553.

[27] Saitou, N. and Nei, M., "The neighbor-joining method: A new method for reconstructing phylogenetic trees", *Mol. Biol. Evol*. *4*, 1987, pp. 406-425.

[28] Tamura, K.., Dudley, J. N. M. and Kumar, S., "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.", *Mol. Biol. Evol. 24*, 2007, pp.1596-1599.

[29] Verma, N. K. and Singh, B. Insight from the structural molecular model of cytidylate kinase from *Mycobacterium tuberculosis. Bioinformation*, *9*, 2013, pp. 680–684.

[30] Bansal, H. Srivastava, S., Chaurasia, A. and Jabalia, N., "A Comparative Study of Antifreeze Proteins from Antarctomyces psychrotrophicus and Typhula ishikariensis using Computational Tools and Servers", *VIVECHAN Int. J. Res*., *5*, 2014, pp. 21-28.

[31] Bansal, H., Narang, D. and Jabalia, N. "Computational characterization of antifreeze proteins of Typhula ishikariensis – Gray Snow Mould",*. J. Proteins Proteomics*, 5, 2014, pp. 169-179.