JPP JOURNAL OF PROTEINS AND PROTEOMICS
6(3), 2015, pp. 245-253

Research Communication

IN SILICO INVESTIGATION OF CYSTEINE PROTEASES FROM *ZINGIBER OFFICINALE*

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

[1]*Amity Institute of Biotechnology, Amity University, Noida 201313, Uttar Pradesh, India*
[2]*Guru Nanak Dev University, Amritsar 143001, Punjab, India*

Abstract: *Zingiber officinale* (Ginger) rhizome is frequently used for the treatment of diarrhea, cold, influenza and as appetite stimulant. Ginger has been reported to be one of the richest sources of proteolytic enzymes. In this study, bioinformatics approach has been adopted to explore properties of all the sequences of cysteine protease. Physico-chemical properties have been studied along with the prediction of motifs for the functional characterization. Apart from these analyses, subcellular location of the sequences, sequence hydropathicity and transmembrane domains have been identified. The secondary structure prediction of the cysteine protease sequences revealed the presence of maximum number of random coils probably due to rich content of more flexible glycine and hydrophobic proline amino acids. The three-dimensional models of the proteins were predicted and evaluated by using some computational approaches. The study may be valuable to understand the structural and functional aspects of the enzymes for academic and industrial proposes.

Keywords: *Zingiber officinale*; cysteine protease; homology modeling

Introduction

Zingiber officinale (Ginger) rhizome is an important spice crop, used for the treatment of diarrhea, cold and as appetite stimulant (Mohammad *et al*., 2012). It is a tuber rhizome used in traditional Indian and Chinese medicines and are effective for wide variety of ailments including respiratory disorders, inflammatory diseases etc. Ginger rhizome has been found to contain proteolytic enzymes (Thompson *et al*., 1973). Proteases are a unique class of enzymes as they possess both degradative and synthetic properties (Jabalia *et al*., 2014). Cysteine proteases have been isolated from ginger (*Zingiber officinale*) rhizome in the past (Choi *et al*., 2000). The meat tenderizing component is ginger protease (GP) or zingipain (Thompson *et al*., 1973; Sakasai *et al*., 1980; Lee *et al*., 1986; Mega *et al*., 1987). GP is highly active against collagen and other connective tissue proteins (Thompson *et al*., 1973). The present market for industrial proteases amounts to about $242 million (US). Two plant proteases (papain and bromelain) account for $17 million (US) sale or 8% market share (Outtrup and Boyce, 1990). Ginger is of considerable interest as a potential new source of protease for industry. GP is one of the least studied of the plant thiol proteases including papain, bromelain, ficin and actinidin (Caygill, 1979; Storey and Wanger, 1986). The objective of this study is to characterize all the sequences of cysteine protease from *Zingiber officinale* 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.

Corresponding Author: **Neetu Jabalia**
E-mail: njabalia@amity.edu

Received: July 17, 2014
Accepted: July 12, 2015
Published: July 17, 2015

Materials and Methods

Sequence analysis - Cysteine protease sequences were retrieved from the SWISS-PROT, a public domain protein database. During the sequence retrieval process, keyword 'cysteine protease *Zingiber officinale'* was used (Table 1). The database search yielded 6 cysteine protease sequences from *Zingiber officinale* which were retrieved in FASTA format and used for further analysis.

Table 1
Cysteine protease sequences from *Zingiber officinale.*

S.No.	*Accession No.*	*Length*	*Description*	*Function*
1	P82474	221	Zingipain-2 (Cysteine proteinase GP-II)	Preferential cleavage of peptides with a proline residue at the P2 position
2	P82473	221	Zingipain-1((Cysteine proteinase GP-I)	Preferential cleavage of peptides with a proline residue at the P2 position.
3	Q5ILG4	466	Cysteine protease gp3b	cysteine-type peptidase activity
4	Q5ILG5	475	Cysteine protease gp3a	cysteine-type peptidase activity
5	Q5ILG7	381	Cysteine protease gp2a	cysteine-type peptidase activity
6	Q5ILG6	379	Cysteine protease gp2b	cysteine-type peptidase activity

Functional characterization - Motifs in the protease sequence were scanned using Motif Search. For sequence hydropathicity and presence of transmembrane, SOSUI server was used (Hirokawa *et al.* 1998).

Physico-chemical characterization - For physico-chemical characterization, isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient (Gill *et al.*, 1989), instability index (Guruprasad *et al*., 1990), aliphatic index (Ikai, 1980) and grand averge hydropathicity (Kyte *et al*., 1982) were computed using the Expasy ProtParam server (Gasteiger *et al.* 2005).

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 protease sequences from *Zingiber officinale* were identified by GORIV.

Tertiary structure prediction - In order to generate three dimensional model, homology modeling approach was applied. The modeling of 3D structure of the sequences was executed by Swiss-Modeler (*http://swissmodel.expasy.org/*) program (Arnold *et al*. 2006; Bordoli *et al*. 2009).

Model visualization and evaluation - For visualization of three dimensional models, the Swiss-PDBViewer was used. Rampage server (*http://mordred.bioc.cam.ac.uk/~rapper/rampage.php*) was used for evaluating and assessing the accuracy of the model (Laskowski *et al*. 1993; Lovell *et al*. 2003).

Protein structure superimposition - The superimposition between three dimensional structure of cysteine protease from *Zingiber officinale* (1CQD), papain from *Carica papaya* (9PAP), Q5ILG4, Q5ILG5, Q5ILG6, Q5ILG7 and P82473 was done by using Schrondinger.

Catalytic triad- To identify putative catalytic triad, multiple sequence alignment was performed by using CLUSTALW (Thompson *et al*., 1994) and conservation of catalytic triad (Aspartic acid (Asp, D), Serine (Ser, S) and Histidine (His, H)) was found.

Results and Discussion

The amino acid sequences of 6 cysteine proteases were retrieved from Swissprot database having

accession numbers Q5ILG4, Q5ILG5, Q5ILG6, Q5ILG7, P82474 and P82473 (Table 1). All the 6 sequences showed cysteine-type peptidase activity but out of these 6 sequences, 2 sequences (P82474 and P82473) showed their catalytic activity with proline residue at P2 position in the N-terminal direction (Schechter and Berger, 1967). The nature of the residue at position P2 of peptide substrates is known to be the most significant in terms of determining specificity for most of the cysteine proteases (Choi *et al*., 2000).

Amino acid composition in the cysteine protease sequences are displayed and compared in Figure 1. A set of conserved amino acid residues located in the vicinity that provide clues to the functions is termed as motif. Motifs were predicted using Motif Search (Table 2). It has been predicted that all the protease sequences have cysteine and histidine as active site residues except P82473, which has asparagine which facilitates the molecular interactions. SOSUI server was used to characterize whether the proteins would be soluble or transmembrane in nature. Protease sequences from *Zingiber officinale* were classified as membrane protein by SOSUI server except P82474 and P82473, which were classified as soluble proteins (Table 3). SOSUI server identified one transmembrane region in Q5ILG4, Q5ILG5, Q5ILG6 and Q5ILG7. Subcellular location of proteases was identified by TargetP and it has been found that Q5ILG4, Q5ILG5, Q5ILG6 and Q5ILG7 are present in secretary pathways (Table 4).

The values of isoelectric point (pI) of the proteases from *Zingiber officinale* were in the range 4.84 to 7.58 indicating that all are acidic except Q5LIG6 (Table 5). The computed isolelectric point will be useful for developing buffer system when these enzymes are to be purified in solution by isoelectric focusing method (Verma *et al*. 2013). The Expasy's Protparam was used to determine the extinction coefficient of the proteases. Extinction coefficient of proteases at 280nm ranged from 43805 to 87540 $\underline{M}^{-1}$ cm^{-1} with respect to the concentration of Cys, Trp and Tyr. The extinction coefficient was high in Q5ILG4 and Q5ILG5 indicating the presence of high concentration of aromatic amino acids. The computed protease concentrations and extinction coefficients will be useful 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 from *Zingier officinale* ranged from 68.78 – 75.38 (Table 5). The very high aliphatic index of these sequences indicated that the cysteine protease will be stable over a wide temperature range (Bansal *et al*, 2014a). 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 (Bansal *et al*, 2014b). GRAVY indices for the sequences ranged from -0.578 to -0.333. This low range of value indicated better interaction with water (Jabalia *et. al*, 2015). Secondary structural features were predicted by GORIV (Figure 2). The results revealed that random coils dominated among secondary structure elements (Bansal *et al*, 2014b). 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.

After the selection of potential template for the enzymes, 3D models were generated by the use of Swiss-modeler program. The PDB ID of 1CQD (*Zingiber officinale*) consisting of 221 amino acids was selected as template from the protein data bank for the rest of the cysteine protease sequences. The proteins displayed very high sequence identities to the template (Table 6). Validation of the three dimensional protein models by using Ramachandran plot in Procheck web-based server revealed that more that 95% residues were in favored regions. The validation parameters for the 3D structures are shown in Table 7. The modeled structures are expected to provide insight into the structural properties of polypeptide chains with significance for

Table 2
Motifs predicted from the protease sequences

Accession No.	*Motif*	*Motif ID*	*Description*	*Start*	*End*
P82474	THIOL_PROTEASE_CYS	PS00139	Eukaryotic thiol (cysteine) proteases cysteine active site.	21	32
	THIOL_PROTEASE_HIS	PS00639	Eukaryotic thiol (cysteine) proteases histidine active site.	159	169
P82473	THIOL_PROTEASE_ASN	PS00640	Eukaryotic thiol (cysteine) proteases asparagine active site	176	195
Q5ILG4	THIOL_PROTEASE_CYS	PS00139	Eukaryotic thiol (cysteine) proteases cysteine active site.	152	163
	THIOL_PROTEASE_HIS	PS00639	Eukaryotic thiol (cysteine) proteases histidine active site.	291	301
	THIOL_PROTEASE_ASN	PS00640	Eukaryotic thiol (cysteine) proteases asparagine active site	308	327
Q5ILG5	THIOL_PROTEASE_CYS	PS00139	Eukaryotic thiol (cysteine) proteases cysteine active site.	161	172
	THIOL_PROTEASE_HIS	PS00639	Eukaryotic thiol (cysteine) proteases histidine active site.	300	310
	THIOL_PROTEASE_ASN	PS00640	Eukaryotic thiol (cysteine) proteases asparagine active site	317	336
Q5ILG7	THIOL_PROTEASE_CYS	PS00139	Eukaryotic thiol (cysteine) proteases cysteine active site.	162	173
	THIOL_PROTEASE_HIS	PS00639	Eukaryotic thiol (cysteine) proteases histidine active site.	300	310
Q5ILG6	THIOL_PROTEASE_CYS	PS00139	Eukaryotic thiol (cysteine) proteases cysteine active site.	160	171
	THIOL_PROTEASE_HIS	PS00639	Eukaryotic thiol (cysteine) proteases histidine active site.	298	308

Table 3
Prediction of hydropathicity and presence of transmembrane domain for the sequences

Accession No.	*Average hydropathicity*	*Protein type*	*Transmembrane region*	*length*	*type*
P82474	-0.5862	Soluble	-	-	-
P82473	-0.4222	Soluble	-	-	-
Q5ILG4	-0.3328	Membrane protein	SFVAVVLLLLLSSSALTAVSAVP	23	Primary
Q5ILG5	-0.3404	Membrane protein	VAVVLLFLLSSSALTAVSVVPPL	23	Primary
Q5ILG7	-0.3772	Membrane protein	TVDNILLLLLLLSSSALAAVSSA	23	Primary
Q5ILG6	-0.4309	Membrane protein	STVDIILLLLFSSSALAAVSSVP	23	Primary

Table 4
Subcellular location of the protease sequences

Name	*Len*	*cTP*	*mTP*	*SP*	*other*	*Loc*	*RC*
P82473	221	0.077	0.132	0.137	0.893	_	2
P82474	221	0.081	0.121	0.09	0.935	_	1
Q5ILG4	466	0.037	0.023	0.973	0.028	S	1
Q5ILG5	475	0.084	0.03	0.963	0.025	S	1
Q5ILG7	381	0.023	0.017	0.987	0.021	S	1
Q5ILG6	379	0.069	0.011	0.971	0.043	S	1

Table 5
Physico-chemical properties of the sequences

Properties	*P82474*	*P82473*	*Q5ILG4*	*Q5ILG5*	*Q5ILG7*	*Q5ILG6*
No. of A. A	221	221	466	475	381	379
M.W	23922.3	24244.2	50956.8	52062.1	41926.7	41749.6
pI	4.82	5.12	6.23	6.17	5.78	7.58
"-" charged residues	24	27	47	48	42	40
"+" charged residues	16	20	43	44	37	41
Extinction coefficients	43805	48275	82040	87540	63745	59735
Instability index	32.22	27.7	33.4	32.61	33.29	30.67
Aliphatic index	68.78	66.15	73.45	73.31	78.85	75.38
Gravy	-0.422	-0.578	-0.333	-0.34	-0.377	-0.431

Table 6
Tertiary structure prediction of protease sequences

Accession No.	*Modelled residue range*	*Template*	*Sequence identity (%)*	*E-value*	*Total energy*
P82473	3-218	1CQD_A	79.17	8.79E-93	-11717
Q5ILG4	134-350	1CQD_A	72.35	2.62E-88	-10890
Q5ILG5	143-358	1CQD_A	74.07	9.38E-90	-11412
Q5ILG6	142-357	1CQD_A	96.3	1.08e-120	-12608
Q5ILG7	144-359	1CQD_A	99.07	3.61e-125	-12719

Table 7
Validation of parameters computed for the modeled 3D structures

Target	*PDB EXPDTA*	*Rampage percentage of residue favored region*	*RMSD (Å)*	*Quality of the model*
P82473	Theoretical Model	96%	0.059	Fairly good model
Q5ILG4	Theoretical Model	96.30%	2.775	Extremely good model
Q5ILG5	Theoretical Model	95%	2.761	Extremely good model
Q5ILG6	Theoretical Model	96.30%	0.057	Fairly good model
Q5ILG7	Theoretical Model	96.30%	0.057	Fairly good model

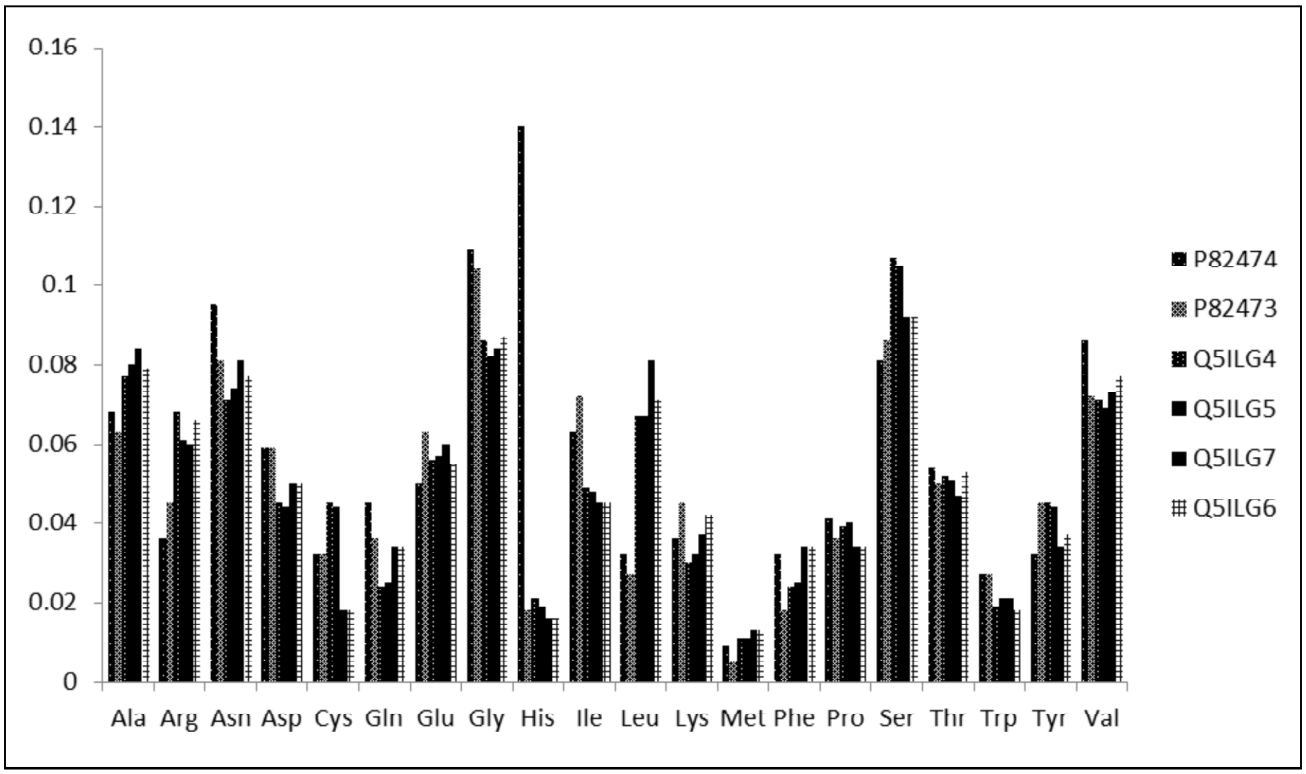

Figure 1: **Graphical representation of the amino acid composition of the protease sequences.**

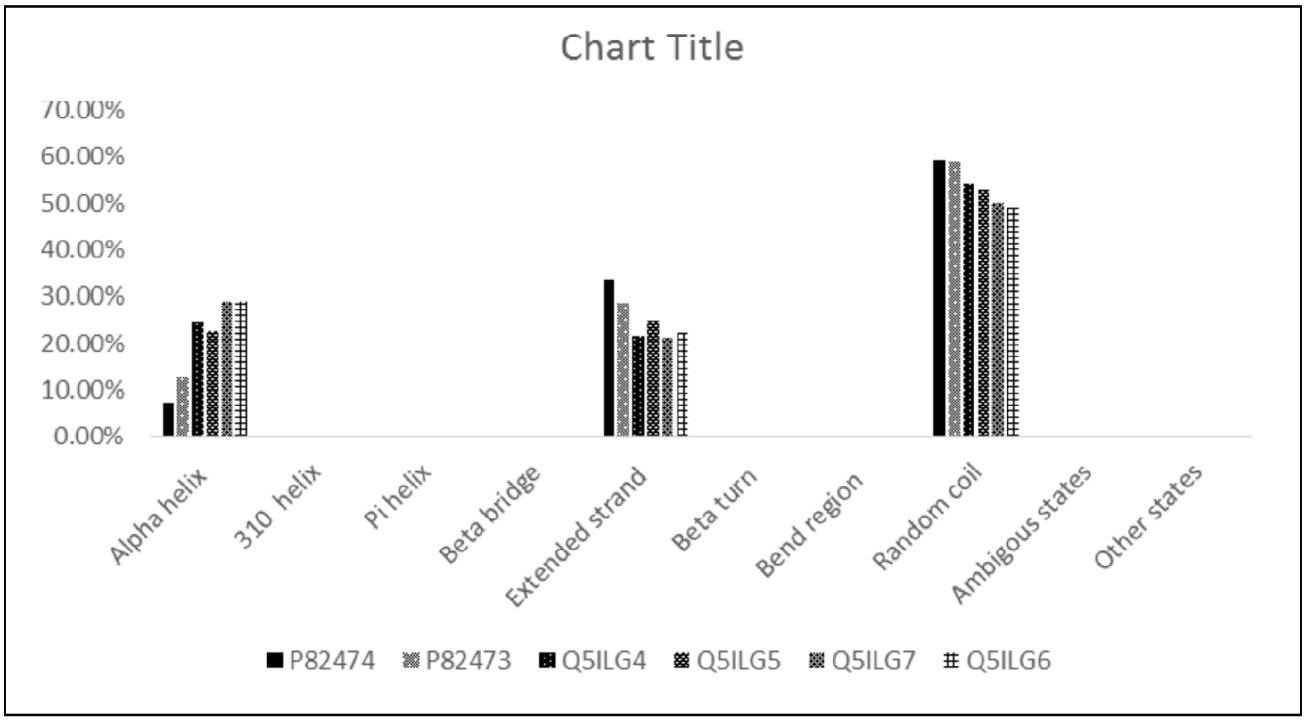

Figure 2: **Graphical representation of the content of secondary structure elements of the protease sequences.**

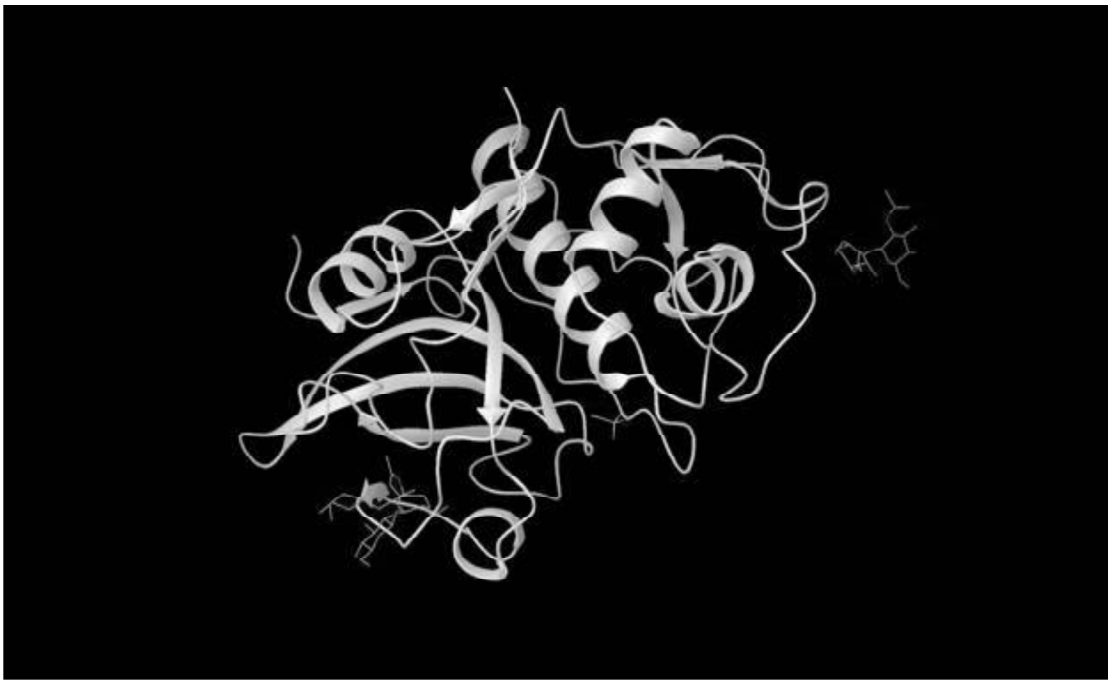

Figure 3: **Ribbon representation of superimposed structures of cysteine protease from *Zingiber officinale* (pdb code: 1CQD), 9PAP, P82473, Q5ILG4, Q5ILG5, Q5ILG6 and Q5ILG7. The codes are detailed in Table 1.**

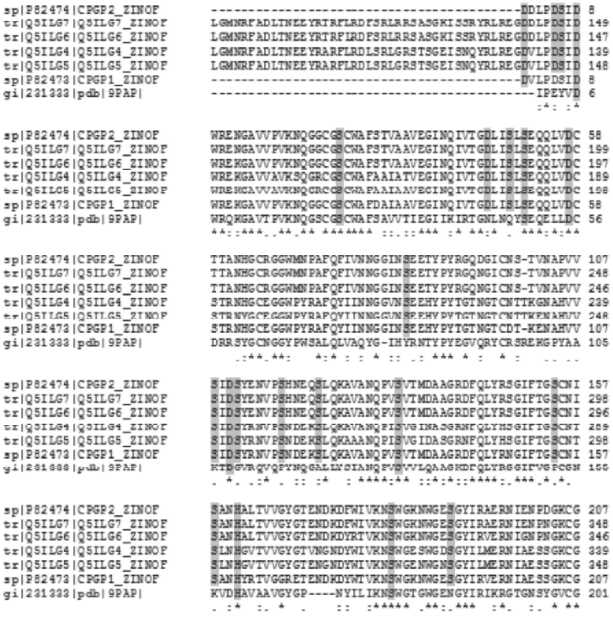

Figure 4: **Multiple sequence alignment of the protease sequences of accession codes 1CQD, 9PAP, P82473, Q5ILG4, Q5ILG5, Q5ILG6 and Q5ILG7 using CLUSTALW. The highlighted regions indicate the conserved putative catalytic residues Asp (D), Ser (S) and His (H).**

understanding protein folding and for molecular design. The superimposition of the three-dimensional structures was performed between cysteine protease from *Zingiber officinale* (1CQD) as a template with papain from *Carica papaya* (9PAP), Q5ILG4, Q5ILG5, Q5ILG6, Q5ILG7 and P82473 using Schrodinger (Figure 3). Root-mean-square deviation (RMSD) value sof the superimposed structures Q5ILG4 and the template 1CQD showed that both of these theoretical models were extremely good. The mutiple sequence alignment confirmed sequence conservation of the catalytic triad residues of Asp, Ser and His (Laskar *et al*., 2011 and Suzuki *et al*., 2014) in all the protease sequences (Figure 4). Thus, these modeled structures were similar to the well studied cysteine protease papain. It is thus expected that the proteases from ginger will have functional properties similar to papain. Papain is the most widely used industrial protease. The proteases from ginger seem to have the potential to be used as industrial enzymes as well, along with papain or as their replacement.

Conclusion

In our study, we have characterized amino acid sequences of multiple cysteine proteases present in *Zingiber officinale*. Total six sequences of cysteine protease have been analyzed to acquire an understanding about their functional properties, physico-chemical properties and various protein structure levels by using *in silico* techniques. 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 structural data is expected to inspire experimental efforts in this area; specifically, the critical assessment of our computational approach can be readily tested for their biochemical relevance. This study will provide understanding about the physicochemical properties and function of cysteine protease in *Zingiber officinale* which will further aid in formulating their uses in academics and industries. It is evident that the proteases are expected to behave like papain in most aspects and might complement papain as industrial proteases.

Acknowledgement

We are grateful to Dr. Chanderdeep Tandon, Head of Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida for his constant support and encouragement during this study.

Abbreviations

Cys, cysteine; GOR, Garnier Osguthorpe Robson; GP, ginger protease; GRAVY, Grand Average hydropathy; O-GlcNAc, O-linked N-acetylglucosamine; OGT, O-linked N-acetylglucosamine transferase; pI, isoelectric point; RMSD, root-mean-square deviation; Trp, tryptophan; Tyr, tyrosine.

References

Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. Bioinformatics. *22,* 195-201.

Bansal, H. Srivastava, S., Chaurasia, A. and Jabalia, N. (2014a). A Comparative Study of Antifreeze Proteins from *Antarctomyces psychrotrophicus* and *Typhula ishikariensis* using Computational Tools and Servers. Vivechan Int. J. Res., *5*, 21-28.

Bansal, H., Narang, D. and Jabalia, N. (2014b). Computational characterization of antifreeze proteins of Typhula ishikariensis – Gray Snow Mould. J. Proteins Proteomics 5, 169-179.

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. and Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. Nat Protoc. *4*, 1-13.

Caygill, J. C. (1979). Sulfhydryl plant proteases. Enzymes Microbiol. Technol. *1*, 233-242.

Choi, K., H. and Laursen, R. A. (2000). Amino-acid sequence and glycan structures of cysteine proteases with proline specificity from ginger rhizome *Zingiber officinale*. Eur. J. Biochem. 267, 1516-1526.

Gasteiger, E. and Walker, J.M. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *In* The Proteomics Protocols Handbook (ed. Walker, J.M.), Humana Press, New Jersey, USA, pp 571-607.

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

Guruprasad, K. (1990). 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*, 155-161.

Hirokawa, T. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics. *14*, 378-379.

Ikai, A.J. (1980). Thermostability and aliphatic index of globular proteins J. Biochem. *88*, 1895-1898.

Jabalia, N., Mishra, P.C. and Chaudhary, N. (2014). Applications, Challenges and Future Prospects of Proteases: An Overview. J. Agro. Nat. Res. Manag. *1*, 179-183.

Jabalia, N., Bansal, H., Mishra, P.C. and Chaudhary, N. (2015). *In silico* comparative analysis of papain family cysteine protease using computational tools and servers. Int. J. Basic and Appl. Eng. Res.. 2, 310-314.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. *157*, 105-132.

Lee, Y.B., Sehnert, D.J. and Ashmore, C.R., (1986). Tenderization of meat with ginger rhizome protease. J. Food Sci. 51, 1558-1559.

Laskar A., Rodger E. H., Chatterjee A. and Mandal, C. (2011). Modeling and structural analysis of evolutionarily diverse S8 family serine proteases. Bioinformation. *7*, 239-245.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. and. Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl.* Crystallogr. *26*, 283-291.

Lovell, S. C., Davis, I. W. and Arendall, W. B. (2003). Structure validation by Cα geometry: Õ, ψ and Cβ deviation. Proteins. *50*, 437-450.

Maiti, Rajarshi, Domselaar, Gary, H. Van and Wishart, D. S. (2004). SuperPose: a simple server for sophisticated structural superposition. Nuc. Acids Res. *32*, 90-94.

Mega, A., Mitsuhashi, T., Tajima, M. and Arakawa, N. (1987). Effects of ginger protease on muscle collagen. J. Home Econ. Jpn. *38*, 923-6.

Mohammad, S. M. and Hamed H. K. (2012). Ginger (Zingiber officinale): A review. J. Med. *Plants* Res. *6*, 4255-4258.

Verma, N. K. and Singh, B. (2013) Insight from the structural molecular model of cytidylate kinase

from *Mycobacterium tuberculosis*. Bioinformation *9*, 680–684.

Outtrup, K. and Boyce, C.O.L. (1990). Microbial proteinase and biotechnology. In Microbial enzyme biotechnology, 2nd ed. (eds. Forgarty, W.M. and Kelly, C.T.), Elsevier Science, New York, p 227.

Sakasai, N., Hayashi, K., Nobuhara, A. and Yamamoto, J. (1980). Reduction in elasticity of Kamaboko (boiled fish-paste) by some additives containing protease. J. Jpn. Soc. Food Sci. Technol. *7*, 371-6.

Schechter, I. and Berger, A. (1967). On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. 27, 157-162.

Storey, R.D. and Wanger, F.W. (1986). Plant proteases a need for uniformity. Phytochemistry. *25*, 2701-2709.

Thompson, E.H., Wolf, I.D. and Allen, C.E. (1973). Ginger rhizome: a new source of proteolytic enzyme. J. Food Sci. *38*, 652-655.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nuc. Acids Res. *11*, 4673-80.

Suzuki. Y., Sakamoto, Y., Tanaka, N., Okada, H., Morikawa, Y. and Ogasawara, W. (2014). Identification of the Catalytic Triad of Family S46 Exopeptidases, Closely Related to Clan PA Endopeptidases. Sci. rep. *4*, 4292.