NUSANTARA BIOSCIENCE
Vol. 6, No. 1, pp. 102-106
May 2014

ISSN: 2087-3948
E-ISSN: 2087-3956
DOI: 10.13057/nusbiosci/n060116

Short Communication: An insight into protein sequences of PTP-like cysteine phytases

VINOD KUMAR[1,2], SANJEEV AGRAWAL[1,♥]

[1]Department of Biochemistry, College of Basic Sciences and Humanities, G.B. Pant University of Agriculture and Technology, Pantnagar-263145, India. Tel. +91-5944-233310, Fax. +91-5944-233473, ♥email: sanjeevagrawal14@rediffmail.com

[2]Akal School of Biotechnology, Eternal University, Baru Sahib, Sirmour-173101, India.

Manuscript received: 26 April 2014. Revision accepted: 30 April 2014.

Abstract. *Kumar V, Agrawal S. 2014. An insight into protein sequences of PTP-like cysteine phytases. Nusantara Bioscience 6: 102-106.* Protein tyrosine phosphatase like cysteine phytases (CPhy) are novel phytases reported in the ruminant microbial community and suggested to play major role in phytate-phosphorus hydrolysis in animal feed. These phytases are very promising to be used in animal feed applications for monogastric animals. Present study deals with utilization of sequence information of 40 CPhy reference protein sequences for their sequential characterization for conserved regions, phylogenetic relationship, biochemical features, superfamily and functional motifs therein. The study reveals that CPhy, not well characterized class of phytases, contains conserved sequence feature which may be important catalytic residues. Five major clusters observed in phylogenetic tree with *Clostridium* sp. as largest cluster. Reported motifs might be used for diversity and expression analysis of CPhy enzymes.

Key words. Cysteine phytase, *in silico* analysis, motifs, phylogenetic tree, phytic acid

INTRODUCTION

Phytases (*myo*-inositol 1,2,3,4,5,6-hexa*kis*phosphate phosphohydrolase) are a special group of phosphatases which catalyzes the stepwise removal of phosphates from phytic acid (*myo*-inositol 1,2,3,4,5,6-hexa*kis*phosphate; IP_6) or its salt phytate (Lei et al. 2013). It is, therefore, useful in various applications, e.g. alleviating antinutritional effect of phytate in animal feed, increase bioavailability of micronutrients to monogastric animals, management of environmental phosphorus pollution (Sapna et al. 2013) and promising for aquafeed application (Kumar et al. 2014b).

Phytases are widely distributed among plants and microbial cells (Hegeman and Grabau 2001; Kumar et al. 2013; Singh et al. 2014; Lei et al. 2013). To develop a suitable phytase for above applications and better understand the catalytic mechanism of diverse groups of phytases, large number of such organisms has been studied in detail for their phytase gene sequences and biochemical properties. Based on the specific consensus sequence, catalytic mechanism and three dimensional structures, so far phytases are therefore classified in four classes, i.e. histidine acid phosphatase (HAPhy), cysteine phytase (CPhy), purple acid phosphatase (PAPhy) and beta-propeller phytase (BPPhy) (Lei et al. 2007; Mullaney and Ullah 2007; Lei et al. 2013). Further classification based on the site of phytic acid dephosphorylation reveals three groups of phytases i.e. 3-phytase (alternative name, 1-phytase; EC 3.1.3.8), 4-phytase (alternative name, 6-phytase; EC 3.1.3.26), and 5-phytase (EC 3.1.3.72) (Kumar et al. 2014a).

Although tremendous work has been carried out related to phytase research, new technologies like sequencing advances, enzyme engineering, proteomics and related bioinformatic studies has further given a new life to this old enzyme (Lei et al. 2013). In addition to this, *in silico* characterization of protein sequences of HAPhy and BPPhy class of phytases has been reported recently (Kumar et al. 2012; Kumar et al. 2014a). These studies have been suggested to be useful in further genetic engineering and classification in important groups of phytases. The use and bioinformatic analysis of resulting DNA and protein sequence information from different studies make it possible to get important predictions and help in the design and success of further studies.

Among the different class of phytases, CPhy class is least studied for its biochemical and important sequence catalytic features. Several CPhy have been reported by rumen bacterial isolates, including *Megasphaera elsdenii, Clostridium perfringens,* and *Clostridium botulinum* (Yanke et al. 1999). Biochemical characteristic of CPhy has not been studied in detail and very little literature is available on characterization of CPhy. In one such study, CPhy from *Selenomonas ruminantium* has been studied and characterized extensively by Puhl et al. (2007). In the present study, the 40 reference protein sequences of CPhy from protein databases were retrieved and analyzed '*in silico*' for their biochemical features, multiple sequence alignment and identity search, phylogenetic tree construction, distribution of motifs and superfamily using various bioinformatics tools.

Materials and methods

The 40 reference protein sequences representing CPhy from NCBI protein database (http: //www.ncbi.nlm.nih.gov) were retrieved in FASTA format for use in this study (Table 1). The protein sequences, which were shown to exhibit phytase activities, were selected for *in silico* study. Their characterization for homology, phylogenetic relationship, functional domain and other biochemical properties was carried out using freely available bioinformatic tools following the methodology of Kumar et al. (2012). For domain search, the Pfam site (http: //www.sanger.ac.uk/software/pfam/search.html) was used. Functional domains and motifs analysis was done using MEME (http: //meme.sdsc.edu/meme/meme.html).

Result and discussion

The 40 reference sequences representing CPhy with GenBank accession number are listed in Table 1. Two conserved regions 'DLR[E/Q]E[S/T]HG[F/L]' and 'WxHFHCxxGxGRT' were obtained in all representative sequences, when analyzed by multiple sequence alignment using ClustalW by MEGA5 (Tamura et al. 2011). High distinctiveness was observed in cysteine phytase protein sequences during alignment. The primary sequence of this enzyme contains a PTP-like signature sequence ($C(X)_5R$), which is ubiquitous among members of the PTP superfamily (Zhang et al. 2002). All PTPs have a phosphate-binding loop (P-loop) at the base of their active site which contains the characteristic PTP signature sequence $C(X)_5R$ (Denu and Dixon 1998; Zhang 2003). Site-directed mutagenesis studies have determined that the cysteine residue present in the P-loop is absolutely required for PTP activity (Puhl et al. 2007). It is a strong nucleophile, and is easily modified by thiol reagents (Sechi and Chait 1998). Chemical modification experiments with alkylating agents also indicate that the P-loop cysteine is required for PTP activity (Zhou et al. 1994). PTPs use the nucleophilic cysteine residue to bind the phosphate monoester of the substrate, forming a thiol-phosphate intermediate (Pannifer et al. 1998). *S. ruminantium* phytase neither contains the conserved RHGXRXP motif nor is affected by divalent metal ions. The active site is located near a conserved cysteine-containing (Cys241) P loop (Chu et al. 2004).

A total of 5 clusters were observed in phylogenetic tree constructed by the Neighbor-Joining method. The largest cluster '1' contains sequences from *Clostridium* sp. (21 sequences). Cluster '2' was composed of *Protochlamydia* (YP_008827.1) and *Parachlamydia acanthamoebae* (ZP_06300753.1). Cluster '3' also contains 2 sequences from *Desulfovibrio* sp. (YP_002953065.1, ZP_07334842.1). Cluster '4' consists of 13 sequences with majority from *Selenomonas* sp. and *Acidaminococcus* sp. Cluster '5' consisted by *Mitsuokella multacida* (ZP_05405390.2) and *Bdellovibrio bacteriovorus* (NP_968118.1) (Figure 1.A and 1.B).

The variations in biochemical features of representing CPhy protein sequences are given in Table 2. The length of protein sequences was found in the range of 283-347 amino acid residues, except 5 sequences from *Clostridium* sp. were 820 amino acid residues long. The theoretical pI value of CPhy sequences was observed to be highest among four classes of phytase and was in the range of 7-10. The instability index is used to measure *in vivo* half life of a protein (Guruprasad et al. 1990). The instability index of 7 CPhy protein sequences was above 40 indicating their low *in vivo* stability (Table 2), while the rest of the sequences with their instability index value below 40 have *in vivo* stability of more than 16 h (Rogers et al. 1986). Aliphatic index of protein measure the relative volume occupied by aliphatic side chains of the amino acids: alanine, valine, leucine and isoleucine. Globular proteins with high aliphatic index have high thermostability and an increase in

Table 1. List of source organism of retrieved CPhy protein sequences (with accession number)

Source organism	Accession no.	Total sequences
Acidaminococcus fermentans	YP_003399467.1	1
Acidaminococcus intestine	YP_004897589.1	1
Acidaminococcus sp.	ZP_03929107.1	1
Bdellovibrio bacteriovorus	NP_968118.1	1
Centipeda periodontii	ZP_08501473.1	1
Clostridium acetobutylicum	NP_149178.1	1
Clostridium botulinum	YP_001787593.1,ZP_02951610.1, YP_002863193.1, ZP_02614565.1, YP_001391515.1, YP_001781827.1, YP_001254710.1	7
Clostridium kluyveri	YP_001394001.1	1
Clostridium ljungdahlii	YP_003781358.1, YP_003781970.1	2
Clostridium perfringens	YP_696211.1, ZP_02633371.1, NP_562440.1, ZP_02635029.1, ZP_02952453.1, ZP_02642534.1, ZP_02863824.1	7
Clostridium sp.	ZP_09204362.1	1
Clostridium sporogenes	ZP_02995608.1	1
Clostridium tetani	NP_782216.1	1
Desulfovibrio fructosovorans	ZP_07334842.1	1
Desulfovibrio magneticus	YP_002953065.1	1
Dialister invisus	ZP_05734150.1	1
Megamonas funiformis	ZP_09733511.1	1
Megasphaera elsdenii	YP_004767129.1	1
Mitsuokella multacida	ZP_09733511.1, ZP_05405390.2	2
Parachlamydia acanthamoebae	ZP_06300753.1	1
Protochlamydia sp.	YP_008827.1	1
Selenomonas flueggei	ZP_04658998.1	1
Selenomonas infelix	ZP_09119488.1	1
Selenomonas noxia	ZP_06603000.1	1
Selenomonas sp.	ZP_07397197.1	1
Selenomonas sputigena	ZP_05898176.1	1

Table 2. Biochemical characteristics of CPhy protein sequences determined by ProtParam server

Accession number	Source organisms	No. of amino acids	Molecular weight	Theoretical pI	Instability index	Aliphatic index
YP_003399467.1	*Acidaminococcus fermentans*	302	34098.8	9.3	31.84	79.26
YP_004897589.1	*Acidaminococcus intestini*	326	36785.4	9.57	37.13	80.25
ZP_03929107.1	*Acidaminococcus* sp.	332	37381.2	9.62	37.34	82.02
NP_968118.1	*Bdellovibrio bacteriovorus*	293	33549.3	8.65	43.78	77.2
ZP_08501473.1	*Centipeda periodontii*	328	37260.4	7.72	29.01	71.77
NP_149178.1	*Clostridium acetobutylicum*	319	36199.5	9.68	27.79	82.51
YP_001787593.1	*Clostridium botulinum*	820	94112.9	7.32	39.07	79.26
YP_002863193.1	*Clostridium botulinum*	820	93906.9	8.67	37.87	81.26
ZP_02614565.1	*Clostridium botulinum*	820	93926	8.88	39.86	79.84
YP_001391515.1	*Clostridium botulinum*	820	93922.8	8.78	39.08	78.88
YP_001781827.1	*Clostridium botulinum*	820	93882.9	8.72	39.73	80.2
YP_001254710.1	*Clostridium botulinum*	820	94022.1	8.83	39.34	80.44
ZP_02951610.1	*Clostridium butyricum*	309	35527	5.35	44.46	84.24
YP_001394001.1	*Clostridium kluyveri*	312	36206.7	9.01	40.77	76.15
YP_003781358.1	*Clostridium ljungdahlii*	343	40226.4	9.21	39.58	84.64
YP_003781970.1	*Clostridium ljungdahlii*	316	36271.8	8.96	38.07	81.65
YP_696211.1	*Clostridium perfringens*	308	35643	8.74	23.97	94.87
ZP_02633371.1	*Clostridium perfringens*	308	35662.9	7.76	24.37	94.25
NP_562440.1	*Clostridium perfringens*	308	35628.9	7.76	24.37	95.32
ZP_02635029.1	*Clostridium perfringens*	308	35610.9	7.76	22.85	96.79
ZP_02952453.1	*Clostridium perfringens*	308	35540.8	7.76	22.45	95.19
ZP_02642534.1	*Clostridium perfringens*	308	35555.8	7.76	22.85	95.84
ZP_02863824.1	*Clostridium perfringens*	308	35554.8	7.76	22.45	95.19
ZP_09204362.1	*Clostridium* sp.	820	93753.2	8.09	32.57	82.96
ZP_02995608.1	*Clostridium sporogenes*	820	93522.1	8.27	41.85	81.5
NP_782216.1	*Clostridium tetani*	307	35509.7	9.16	29.91	89.8
ZP_07334842.1	*Desulfovibrio fructosovorans*	283	31305.2	5.73	46.2	82.79
YP_002953065.1	*Desulfovibrio magneticus*	331	35004	10.19	44.77	88.34
ZP_05734150.1	*Dialister invisus*	408	46674.4	9.1	38.51	76.08
ZP_09733511.1	*Megamonas funiformis*	341	39558.9	8.21	28.77	81.55
YP_004767129.1	*Megasphaera elsdenii*	347	39486.4	4.79	42.89	69.14
ZP_05405389.1	*Mitsuokella multacida*	640	73136.4	9.06	29.94	74.22
ZP_05405390.2	*Mitsuokella multacida*	323	36779.8	9.16	35.9	70.99
ZP_06300753.1	*Parachlamydia acanthamoebae*	320	37249.7	6.46	39.84	90.5
YP_008827.1	*Protochlamydia* sp.	311	35620.6	7.16	38.55	89.29
ZP_04658998.1	*Selenomonas flueggei*	328	36888.7	7.25	28.98	72.38
ZP_09119488.1	*Selenomonas infelix*	328	37258.2	7.75	33.12	69.7
ZP_06603000.1	*Selenomonas noxia*	328	37329.3	8.68	26.82	69.39
ZP_07397197.1	*Selenomonas* sp. oral taxon	328	36818.5	6.99	31.55	70.64
ZP_05898176.1	*Selenomonas sputigena*	334	37515	4.76	38.84	75.48

Table 3. Distribution of Superfamily among CPhy protein sequences determined using superfam server

Motifs	Motif present in no. of sequence	Motif width	Amino acid sequence	Domain
1	40	50	TDHKWPTDEMVDYFVQFVKSMPKDTWLHFHCQAGIGRTTTFMIMYDMMKN	PTPc superfamily
2	40	29	ICIVDLRQESHGFINGYPVSWYGEHNWAN	No putative conserved domains
3	40	29	PNREGLDTLNISGSQQFSPQNLPLLVKSI	No putative conserved domains
4	40	29	PPQTIIPTKVMTEEQLVEHNGMRYVRIPV	No putative conserved domains
5	16	41	ADEIINRQLALAGFDEKHMKSFPNKERHDFFQKFYEYVKEQ	No putative conserved domains
6	8	50	HYVTFIMSDGDNQQWNLGTNYGSPKWYGSPYRGNFNLGWSLSPSLYYLAP	GxGYxYP superfamily
7	8	50	RDKVFSSMDPNSICLGWGPDEFINVSTSSKHGVSMIAADWSYNLTVLSAF	GxGYxYP superfamily
8	8	50	KIPTHLYVISQNKMTSSERTMIATLQGIVNNHCSHQIYTLNSSQPDYQIW	GxGYxYP superfamily
9	8	50	FYNNKLWDKFTVKPNIQGLFYLDYRKHNNYHGEIIWSNNKPIVSCRDLLW	GxGYxYP superfamily
10	8	50	GDCRNTDKDWAYNNLWNSGLNHSIVIQLSPEKETALRDYAIMTKSLIFYE	GxGYxYP superfamily

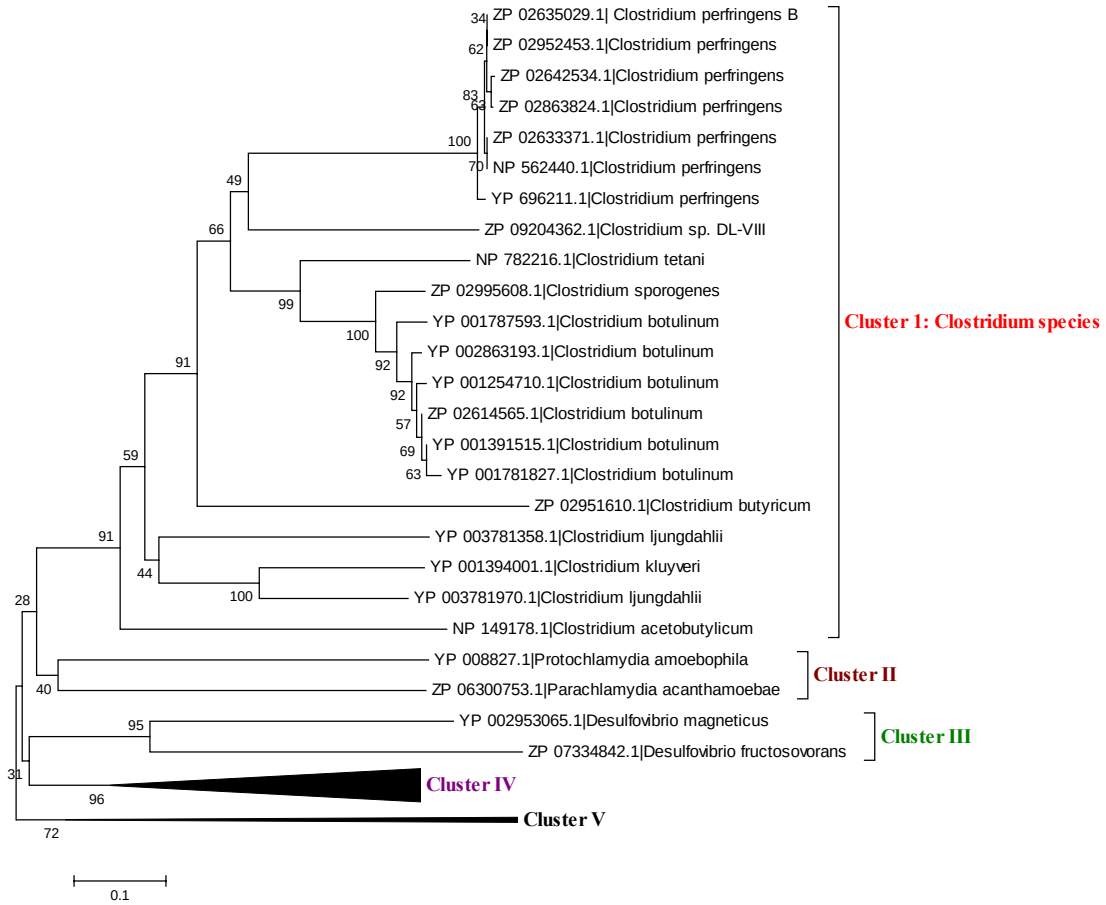

Figure 1.A. Phylogenetic tree constructed by NJ method based on CPhy protein sequences

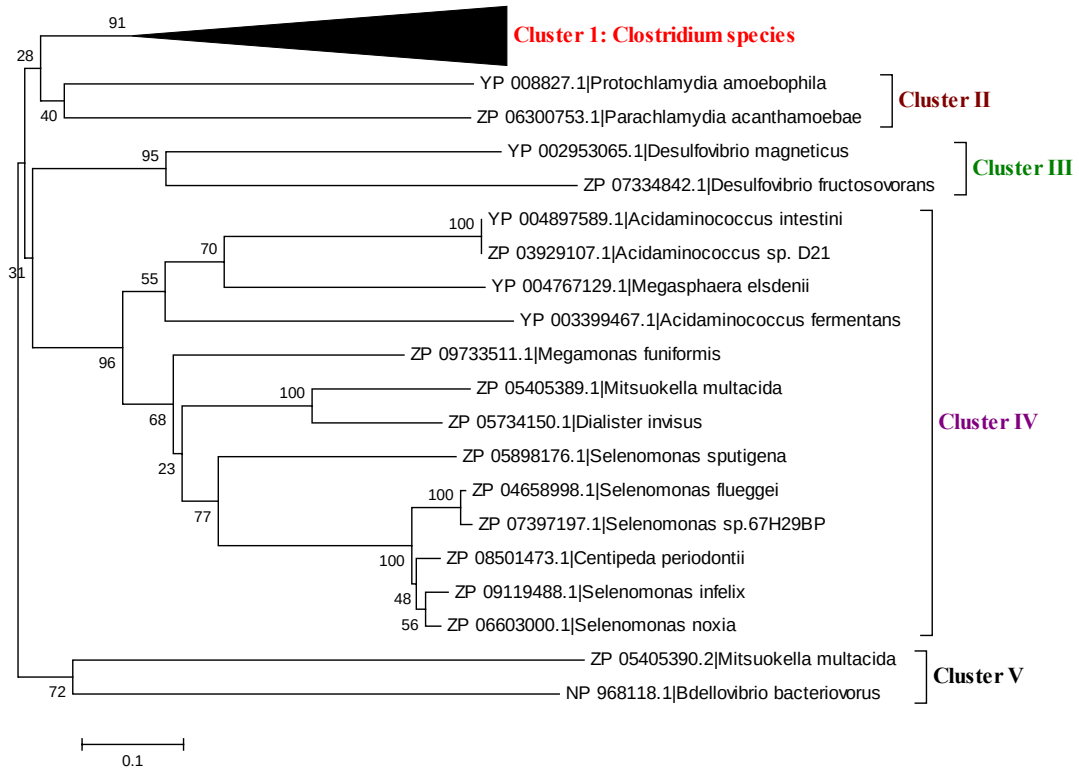

Figure 1.B. Phylogenetic tree constructed by NJ method based on CPhy protein sequences

aliphatic index increases protein thermostability (Atsushi 1980; Rawlings et al. 2006). Aliphatic index of CPhy protein sequences was observed in the range of 70-95, suggesting sequences were varied in their thermostability (Table 2). Superfam analysis revealed the presence of 'phosphotyrosine protein phosphatase II superfamily' and 'myo-inositol hexaphosphate phosphohydrolase PhyA family' in all 40 protein sequences. Protein tyrosine phosphatases (PTP) catalyze the dephosphorylation of phosphotyrosine peptides; they regulate phosphotyrosine levels in signal transduction pathways. The depth of the active site cleft renders the enzyme specific for phosphorylated Tyr (pTyr) residues, instead of pSer or pThr.

Analysis of 10 motifs by MEME suite with provided parameters revealed 50 amino acids long motif '1' 'TDHKWPTDEMVDYFVQFVKSMPKDTWLHFHCQA GIGRTTTFMI MYDMMKN' was present in all cysteine phytase protein sequences. The functional domain found in this motif was similar to PTPc superfamily. This family has a distinctive active site signature motif, HCSAGxGRxG, characterized as either transmembrane, receptor-like or non-transmembrane (soluble) PTPs. Receptor-like PTP domains tend to occur in two copies in the cytoplasmic region of the transmembrane proteins, only one copy may be active. Other motifs (6 to 10) were found similar to GxGYxYP superfamily. This family carries a characteristic sequence motif, GxGYxYP, but is of unknown function. Associated families are sugar-processing domains. Complete list of motifs with their characteristics is given in Table 3.

In conclusion, this *in silico* study for phylogenetic clustering, conserved motifs sequences and biochemical features of phytases from class CPhy, could be key information for their further classification and genetic modification within key sequence features for the development of novel phytase with desired properties. Conserved motif sequences are important for conserved primer design and diversity study thereafter.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Department of Science and Technology for providing infrastructural facility in the form of DST-FIST grant to Department of Biochemistry, G.B. Pant University of Agriculture and Technology, India. Lead author is grateful to Indian Council of Agriculture Research (ICAR), for providing financial assistance in the form of ICAR-SRF.

REFERENCES

Atsushi I. 1980. Thermostability and aliphatic index of globular proteins. J Biochem 88: 1895-1898.

Chu HM, Guo RT, Lin TW, Chou CC, Shr HL, Lai HL, Tang TY, Cheng KJ, Selinger BL, Wang AHJ. 2004. Structures of *Selenomonas ruminantium* Phytase in Complex with Persulfated Phytate: DSP Phytase Fold and Mechanism for Sequential Substrate Hydrolysis. Structure 12: 2015-2024.

Denu JM, Dixon JE. 1998. Protein tyrosine phosphatases: mechanisms of catalysis and regulation. Curr Opin Chem Biol 2: 633-641.

Guruprasad K, Reddy BVB, Pandit MW. 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 Engineer 4: 155-161.

Hegeman CE, Grabau EA. 2001. A novel phytase with sequence similarity to purple acid phosphatases is expressed in cotyledons of germinating soybean seedlings. Plant Physiol. 126: 1598-1608

Kumar V, Sangwan P, Singh G, Verma AK, Agrawal S. 2014a. Cloning, sequencing, expression and in silico analysis of β-propeller phytase from *Bacillus licheniformis* strain PB-13. Biotech Res Internat 2014: 1-11, Article ID 841353, doi: 10.1155/2014/841353

Kumar V, Sangwan P, Verma AK, Agrawal S. 2014b. Molecular and Biochemical Characteristics of Recombinant β-Propeller Phytase from *Bacillus licheniformis* Strain PB-13 with Potential Application in Aquafeed. Appl Biochem Biotechnol DOI 10.1007/s12010-014-0871-9.

Kumar V, Singh G, Verma AK, Agrawal S. 2012. In silico characterization of histidine acid phytase sequences. Enzyme Res 2012: 1-8, 845465, doi: 10.1155/2012/845465

Kumar V, Singh P, Jorquera M, Sangwan P, Kumar P, Verma AK, Agrawal S. 2013. Isolation of phytase-producing bacteria from Himalayan soils and their effect on growth and phosphorus uptake of Indian mustard (*Brassica juncea*). World J Microbiol Biotech 29: 1361-1369.

Lei XG, Weaver JD, Mullaney E, Ullah AH, Azain MJ. 2013. Phytase, a new life for an "old" enzyme. Annu Rev Anim Biosci 1(1): 283-309.

Mullaney EJ, Ullah AH. 2003. The term phytase comprise several different classes of enzymes. Biochem Bioph Res Co 312: 179-184.

Pannifer ADB, Flint AJ, Tonks NK, Barford D. 1998. Visualization of the cysteinyl-phosphate intermediate of a protein-tyrosine phosphatase by X-ray crystallography. J Biol Chem 273: 10454-10462.

Puhl AA, Gruninger RJ, Greiner R, Janzen TW, Mosimann SC, Selinger LB. 2007. Kinetic and structural analysis of a bacterial protein tyrosine phosphatase-like myo-inositol polyphosphatase. Protein Sci 16: 1368-1378.

Rawlings ND, Morton FR, Barrett AJ. 2006. MEROPS: the peptidase database. Nucleic Acids Res 34: 270-272.

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

Sapna, Singh B, Singh D, Sharma KK. 2013. Microbial Phytases in Skirmishing and Management of Environmental Phosphorus Pollution. In: Biotechnology for Environmental Management and Resource Recovery. Springer India.

Sechi S, Chait BT. 1998. Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. Anal Chem 70: 5150-5158.

Singh P, Kumar P, Agrawal S. 2014. Evaluation of phytase producing bacteria for their plant growth promoting activities. Internat J Microbiol 2014: 1-7, Article ID: 426483, doi: 10.1155/2014/426483

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739.

Yanke L, Selinger L, Cheng KJ. 1999. Phytase activity of *Selenomonas ruminantium*: a preliminary characterization. Lett Appl Microbiol 29: 20-25.

Zhang ZY, Zhou B, Xie L. 2002. Modulation of protein kinase signaling by protein phosphatases and inhibitors. Pharmacol Therapeut 93: 307-317.

Zhang ZY. 2003. Mechanistic studies on protein tyrosine phosphatases. Prog Nucl acid Res Mol Biol 73: 171-220.

Zhou MM, Logan TM, Theriault Y, Van Etten RL, Fesik SW. 1994. Backbone 1H, 13C, and 15N assignments and secondary structure of bovine low molecular weight phosphotyrosyl protein phosphatase. Biochem 33: 5221-5229.