ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

In-silico Analysis of Plant Protease Inhibitors and Motif Analysis.

Parameswari P*, and Sivaramakrishna Prasad.

*Department of Bioinformatics, Bharath University, Chennai, Tamil Nadu, India.
Student, Department of Bioinformatics, Bharath university, Chennai, , Tamil Nadu, India.

ABSTRACT

Plant protease inhibitors (PIs) are small proteins presents in plants. PIs' plays major role in the inhibition of proteases present in insects' guts to decrese the digestion, which leads to death of insects. Pests are the major causes which destroys most of the crops. Generally chemical pestisides will be used to control the pest but in recent years PIs playing major role in pest management. As the role of inhibitors is simply achieved by the activation of single genes, several transgenic plants expressing PIs have been produced in the last 15 years and tested for enhanced defensive capacities, with particular efforts against pest insects. In this study we have analyzed the Motifs and Evolutionary relationship using 30 Plant protease inhibitors sequences from various sources. The result found in this study reveals that most of the PIs shares common motif which consist of the amino acid "Cys", "Pro", "Arg", "Lys", "Ser".

Keywords: Proteinase inhibitors, insilico, Motif, aminoacids.

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Plant Proteinase inhibitors (PIs), which play a potent defensive role against predators and pathogens, are natural, defense-related proteins often present in seeds and induced in certain plant tissues by herbivory or wounding. The Proteinase inhibitor IIs (PIN 2) and its molecular biology, including transgenic plants expressing proteinase inhibitors against insect, pests, and pathogens, esp. in lettuce (*Lactuca sativa* L.) and Chinese flowering cabbage (*Brassica campestris* ssp*. parachinensis*), which are widely cultivated [1]. The diverse plant proteinase inhibitors (PIs) genes from different plant species have been isolated and their products with one or more genes are targeted at different biochemical and physiological process with the insect, and well played a potent defensive role against insects and pathogens. The use of recombinant PIs and synergistic activation to protect plants has been incorporated in integrated pest management program. Though they may not replace the use of chemical pesticides in the near future, but effectively complement it [2]. Insect pests are a major cause of damage to the world's commercially important agricultural crops. Current strategies aimed at reducing crop losses rely primarily on chemical pesticides. Alternatively transgenic crops with intrinsic pest resistance offer a promising alternative and continue to be developed. Cysteine proteinases isolated from insect larvae are inhibited by both synthetic and naturally occurring cysteine proteinase inhibitors [3]. The inhibitory activity of Bovine pancreatic trypsin inhibitor (aprotinin), a natural polypeptide and a proteinase inhibitor, was demonstrated on gut proteinases of these lepidopteran borers of sugarcane using commercially available aprotinin Trypsin inhibitor (PPTI) was purified from seeds of the native Brazilian tree Poecilanthe parviflora. The N terminal sequence of PPTI showed high degree of homology with other Kunitz type inhibitor. Trypsin like activity in midguts of larval *Diatraea saccharalis, Anagastra kuehniella, Spodoptera frugiperda. Corcyra cephalonica* were substantially inhibited by PPTI [4]**.** Knowledge on the role of aspartic proteinases in insect digestion is limited than that of cysteine proteinases. In species of six families of the order hemiptera, aspartic proteinases (cathepsin D-like proteinases) were found along with cysteine proteinases [5]. Isolation of the midgut proteinases from the larvae of cowpea weevil, C. maculatus and bruchid Zabrotes subfaceatus [6]. Cystanins have also been characterized from potato and avacado [7]**.**Components of thesesignalling pathways are mostly similar to those implicated in other signalling cascades which include reversible protein phosphorylation steps, calcium/calmodulinregulated events, and production of active oxygen species [8]. *P.melanarius* showed no detectable alterations in mortality, weight gain or food consumption when feeding on *D.reticulatum* [9]. The polypeptide Lawrence, P.K. and Koundal, K.R. 98 activates a lipid-based signal transduction pathway in which linolenic acid, is released from plant membranes and converted into an oxylipin signaling molecule, jasmonic acid [10].

MATERIALS AND METHODS

COLLECTION OF SEQUENCES:

The plant protease inhibitors sequences were collected from NCBI data base, available in the public domain at http://ncbi.nlm.nih.gov/entrez/. NCBI was established in 1988 as a national resource for molecular biology information; NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. Using the search option the sequences were collected. The accession number for the protein sequences are tabulated below (table 1).

Table 1: Input sequences taken from NCBI.

S.NO	PROTEIN SEQUENCE	ACCESSION NO
1	Sinapis alba (white mustard)	AAB24340.1
2	Phaseolus vulgaris	AAB26657.1
3	Phaseolus vulgaris 1	AAB26656.1
4	Pisum sativum (pea)	CAC24566.1
5	Nicotiana glutinosa	AAF18450.1
6	Capsicum annuum	AAX84036.1
7	Solanum copersicum copersicon esculentum)	AAG12170.1
8	Arabidopsis thaliana (thale cress)	NP_175202.1
9	Glycine max (soybean)	CAA48655.1
10	Vigna mungo (black gram)	ABD97865.1

11	Medicago sativa	CAA56254.1
12	Phaseolus grayanus	CAQ17032.1
13	Phaseolus oligospermus	CAO82009.1
14	Phaseolus glabellus	CAM88858.1
15	Ricinus communis (castor bean)	CAA89697.1
16	Solanum tuberosum (potato)	AAM21646.1
17	Brassica rapa (Brassica campestris)	ABK78689.1
18	Phaseolus lunatus (lima bean)	AAK97768.1
19	Vigna unguiculata subsp. cylindrica (horse gram)	AAK97765.1
20	Arachis hypogaea (peanut)	AAP93913.1
21	Phaseolus filiformis	CAL69282.1
22	Medicago truncatula (barrel medic)	AAQ63885.1
23	Brassica napus (rape)	AAM73807.1

Multiple sequence alignment:

The downloaded sequences were clustered into groups using the pairwise alignment score obtained from CLUSTALW. A multiple sequence alignment (MSA) is a sequence alignment of three or more biological sequences, generally protein, DNA, or RNA. In general, the input set of query sequences are assumed to have an evolutionary relationship by which they share a lineage and are descended from a common ancestor. From the resulting MSA, sequence homology can be inferred and phylogenetic analysis can be conducted to assess the sequences' shared evolutionary origins. Visual depictions of the alignment as in the image at right illustrate mutation events such as point mutations (single amino acid or nucleotide changes) that appear as differing characters in a single alignment column, and insertion or deletion mutations (indels or gaps) that appear as hyphens in one or more of the sequences in the alignment. Multiple sequence alignment is often used to assess sequence conservation of protein domains, tertiary and secondary structures, and even individual amino acids or nucleotides.

ClustalW2 is a most widely used multiple sequence alignment computer program. It performs the global multiple sequence alignment. ClustalW calculates gaps in a noval way designed to place them between conserver regions. It also has options for adding one or more additional sequence with weights or an alignment to an existing alignment. The ClustalW program can be accessed using the URL .www.ebi.ac.uk/clustalw/

MOTIF SEARCH

The plant protease inhibitors motifs were discovered using MEME program. The MEME program can be accessed using the URL. http://meme.sdsc.edu/meme/meme-intro.html. MEME (Multiple EM for motif elicitation) is a tool for discovering motifs in a group of related DNA or protein sequences. A motif is a sequence pattern that occurs repeatedly in a group of related protein or DNA sequences. MEME represents motifs as position- dependent letter- probability matrices which describe the probability of each possible letter at each position in the pattern. Individual MEME motifs do not contain gaps. Statistical modeling techniques were used to automatically choose the best width, number of occurrences, and description of each motif. We have submitted the total protein sequence of plant protease to the MEME tool and the results were obtained through e-mail.

PHI BLAST:

The motifs which we have discovered in the MEME tool were submitted to the database similarity tool such as Pattern Hit Initiated BLAST (PHI-BLAST) to identify the similar sequences with the same pattern. PHI BLAST treats two occurrence of the same pattern within the query sequence as two independent sequences. This can be seen in a number of ways, from the statistical analysis at the end of the search results. This tool is only available for database protein searches.

ISSN: 0975-8585

RESULT AND DISCUSSION

CLUSTER ANALYSIS:

The cluster analysis has been done using the pair wise alignment score obtained from CLUSTAL W tool. We found 435 possible pair wise alignment from the CLUSTALW for 30 input sequences. The sequences which has >80% of alignment score were considered as GROUP I (Indicated in Blue colour). The sequences which has >50% of alignment score were considered as GROUP II (Indicated in red colour). The sequences which has >30% of alignment score were considered as GROUP III (Indicated in Brown colour) and the sequences which has <30% of alignment score were considered as GROUP IV (Indicated in Green colour).

CLUSTAL 2.0.10 Multiple Sequence Alignments

Sequence format is Pearson
Sequence 1: gi|260921|gb|AAB24340.1| 63 aa
Sequence 2: gi|300397|gb|AAB26657.1| 85 aa
Sequence 3: gi|300396|gb|AAB26656.1| 85 aa
Sequence 4: gi|12329988|emb|CAC24566.1| 104 aa
Sequence 5: gi|6581087|gb|AAF18450.1|AF205 506 aa
Sequence 6: gi|62467832|gb|AAX84036.1| 204 aa
Sequence 7: gi|12007536|gb|AAG12170.1| 148 aa
Sequence 8: gi|15220298|ref|NP_175202.1| 391 aa
Sequence 9: gi|18541|emb|CAA48655.1| 110 aa
Sequence 10: gi|90762023|gb|ABD97865.1| 113 aa
Sequence 11: gi|509374|emb|CAA56254.1| 113 aa
Sequence 12: gi|169807808|emb|CAQ17032.1| 107 aa
Sequence 13: gi|154147103|emb|CAO82009.1| 107 aa
Sequence 14: gi|145586432|emb|CAM88858.1| 107 aa
Sequence 15: gi|1638842|emb|CAA89697.1| 209 aa
Sequence 16: gi|20386383|gb|AAM21646.1|AF49 221 aa
Sequence 17: gi|118197452|gb|ABK78689.1| 202 aa
Sequence 18: gi|15529129|gb|AAK97768.1| 104 aa
Sequence 19: gi|15529123|gb|AAK97765.1| 103 aa
Sequence 20: gi|33090235|gb|AAP93913.1| 80 aa
Sequence 21: gi|118026422|emb|CAL69282.1| 120 aa

The Plant Protease Inhibitors Phaseolus lunatus- lima bean (AAK97768.1), Vigna unguiculata subsp cylindrica -horse gram (AAK97765.1), Phaseolus filiformis (CAL69282.1), Lablab purpureus -Indian field bean (AAK97770.1), Phaseolus vulgaris 2 (AAK97767.1) are closely related to Vigna mungo-black gram (ABD97865.1). Likewise Phaseolus oligospermus (CAO82009.1), Phaseolus glabellus (CAM88858.1), Vigna unguiculata-cowpea (AAS79232.1) are closely related to Phaseolus oligospermus (CAO82009.1). Phaseolus glabellus (CAM88858.1), Vigna unguiculata-cowpea (AAS79232.1) are closely related to Phaseolus oligospermus (CAO82009.1). Vigna unguiculata subsp cylindrica -horse gram (AAK97765.1), Phaseolus filiformis (CAL69282.1), Lablab purpureus -Indian field bean (AAK97770.1) and Phaseolus vulgaris 2 (AAK97767.1) are closely related to Phaseolus lunatus- lima bean (AAK97768.1), Then the protease Phaseolus filiformis (CAL69282.1), Lablab purpureus -Indian field bean (AAK97770.1) and Phaseolus vulgaris 2 (AAK97767.1) are closely related to Vigna unguiculata subsp cylindrica -horse gram (AAK97765.1). Lablab purpureus -Indian field bean (AAK97770.1) and Phaseolus vulgaris 2 (AAK97767.1) are closely related to Phaseolus filiformis (CAL69282.1) respectively with the alignment score of >80% (GROUP I). Hence the above mentioned combinations are Homologous to each other.

The Plant Protease Inhibitors Pisum sativum -pea(CAC24566.1) , Medicago sativa (CAA56254.1), Phaseolus grayanus (CAQ17032.1) and Vigna unguiculata-cowpea (AAS79232.1) are related to Phaseolus vulgaris(AAB26657.1) and Phaseolus vulgaris(AAB26657.1). Likewise Musa acuminata AAA Group (ABL63911.1) and Oryza sativa Indica Group (CAB88209.1) are related to Phaseolus lunatus- lima bean (AAK97768.1) and Musa acuminata AAA Group (ABL63911.1) and Oryza sativa Indica Group (CAB88209.1) are related to Phaseolus vulgaris 2 (AAK97767.1). The protease inhibitors Nicotiana glutinosa(AF205851_1) and Solanum

ISSN: 0975-8585

copersicum copersicon esculentum)(AAG12170.1), Medicago truncatula -barrel medic (AAQ63885.1) and Brassica napus –rape (AM73807.1), Vigna unguiculata-cowpea (AAS79232.1) and Oryza sativa Indica Group (CAB88209.1), Lablab purpureus -Indian field bean (AAK97770.1) and Musa acuminata AAA Group (ABL63911.1) are related respectively with the alignment score of >30% (GROUP III). Hence the above mentioned combinations are may Similar to each other.The Phylogenetic tree was constructed by using the above mentioned pair wise alignment score (Figure 1). The rooted tree revelas that GROUP I sequences were shares a common ancesteral node with minimum branch lenth. The GROUP II, GROUP III and GROUP IV clusters are connected with each other with varying branch length. But all the Groups were connected with common out group. Thus from this tree we can conclude that all the sequences which we have taken as query may have common ancestor.

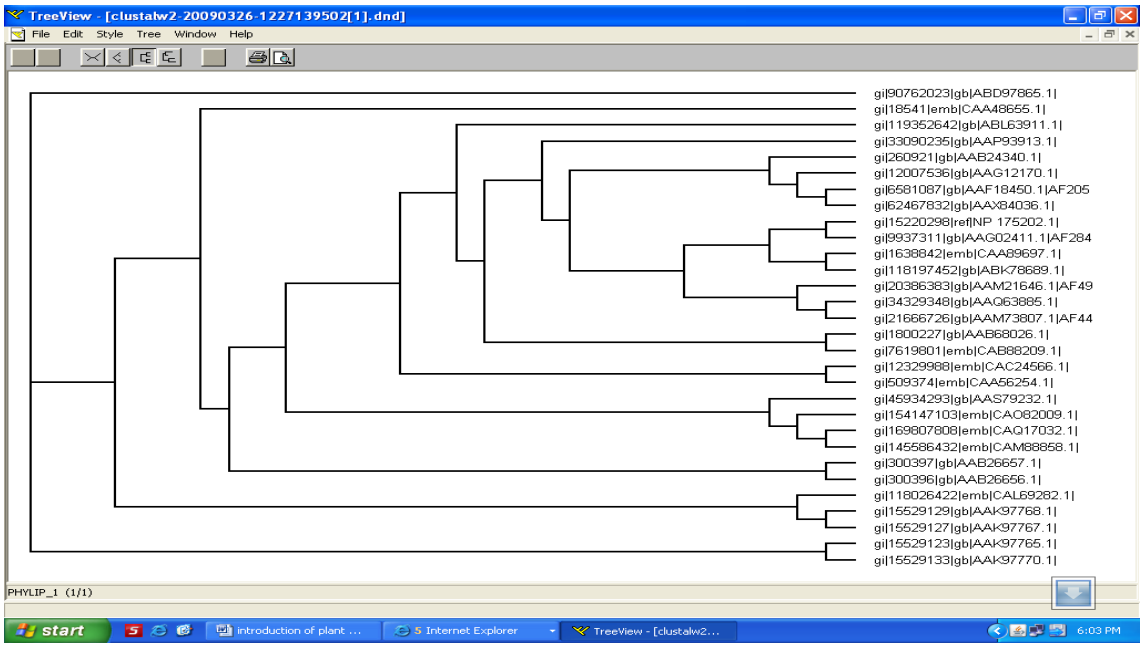

Figure 1: Graphical representation of pairwise alignment score obtained from ClustalW (Rooted tree)

MOTIF ANALYSIS:

Three motifs were discovered from the data set which we have uploaded as query to the MEME tool. Motif 1 has the width of 11 sites and E-value of 1.8e-079. We have listed the Motif alignment in the following table with P-Value and the Start site of the Motif in the sequence (Table 2, Figure 2). Motif regions were coloured based on the biochemical properties of the amino acids. Most Hydrophobic amino acids (**A, C, F, I, L, V, W** and **M**) indicated in blue colour, Polar, non-charged, non-aliphatic residues (**N,Q,S,T**) are indicated in Green colour , the Acidic amino acids (**D,E**) are indicated in Megenta colour, Positively charged amino acids (**K,R**) are indicated in red colour, Histidine , Glycine, Proline and Tyrosine are indicated in Pink, Orange, Brown and Turquoise colour respectively.

The Motif 1 reveals that the query sequences AAK97767.1, CAL69282.1, AAK97765.1, AAK97768.1, AAB26656.1 and AAB26657.1 has the pattern consist of "**CTKSIPPQCRC"** amino acid in the respective order**.** In the pattern of CAC24566.1 amino acid **K**. In the above mentioned pattern has been replaced with **T** (non-charged) **Q** (non charged) **("CTKSTPPQCQC ")** amino acids in the position fifth and eighth and tenth respectinvely. In the next pattern of CAM88858.1 amino acids K (positively charged), **T, P, Q** and **Q (**positively charged**)** present in the third, fifth, seventh, eighth and tenth position respectively in the above mentioned pattern has been replaced with **R (**positively charged**), I**(Hydrophobic), **G, T** (non-charged) and **R (**positively charged**) ("CTRSIPGTCRC")** amino acids in the position third, fifth, seventh, eighth and tenth respectinvely. In the next pattern of CAQ17032.1 amino acids **I** (Hydrophobic) present in the fifth position in the above mentioned pattern has been replaced with **M** (Hydrophobic) **("CTRSMPGTCRC ")** amino acids in the position of third respectively.

ISSN: 0975-8585

Table 2: MEME results of motif

NAME	START	P-VALUE	SITES
gi|15529127|gb|AAK97767.1|	54	2.38e-14	SSKACCDHCACTKSIPPQCRCSDLRLNSCHS
gi|118026422|emb|CAL69282.1|	61	2.38e-14	SSKACCDQCACTKSIPPQCRCSDLRLNSCHS
gi|15529123|gb|AAK97765.1|	53	2.38e-14	SSKPCCDQCACTKSIPPQCRCTDVRLNSCHS
gi|15529129|gb|AAK97768.1|	54	2.38e-14	SSKPCCDHCACTKSIPPQCRCTDLRLDSCHS
gi|300396|gb|AAB26656.1|	27	2.38e-14	SSKPCCBHCACTKSIPPQCRCSBLRLNSCHS
gi|300397|gb|AAB26657.1|	26	2.38e-14	SSKPCCBHCACTKSIPPQCRCSBLRLNSCHS
gi|12329988|emb|CAC24566.1|	55	1.34e-13	SNKACCDSCLCTRSIPPQCQCNDIGETCHSA
gi|15529133|gb|AAK97770.1|	53	2.32e-13	SSKPCCDHCACTKSIPPQCHCSDLRLNSCHS
gi|45934293|gb|AAS79232.1|	50	2.32e-13	SSEPCCDSCICTKSIPPQCHCTDIRLNSCHS
gi|18541|emb|CAA48655.1|	52	2.32e-13	SSKPCCDQCACTKSNPPQCRCSDMRLNSCHS

Motif 2 has the width of 15 sites and E-value of 7.9e-067. We have listed the Motif alignment in the following table with P-Value and the Start site of the Motif in the sequence (Table 3 , Figure 3). Motif regions were coloured based on the biochemical properties of the amino acids. Most Hydrophobic amino acids (**A, C, F, I, L, V, W** and **M**) indicated in blue colour, Polar, non-charged, non-aliphatic residues (**N,Q,S,T**) are indicated in Green colour , the Acidic amino acids (**D,E**) are indicated in Megenta colour, Positively charged amino acids (**K,R**) are indicated in red colour, Histidine , Glycine, Proline and Tyrosine are indicated in Pink, Orange, Brown and Turquoise colour respectively.

Table 3: MEME results of motif 2

NAME	START	P-VALUE	SITES
gi|118026422|emb|CAL69282.1|	73	1.65e-18	KSIPPQCRCSDLRLNSCHSACKSCICTFSIPAQCV
gi|15529127|gb|AAK97767.1|	66	4.53e-18	KSIPPQCRCSDLRLNSCHSECKSCICTLSIPAQCV
gi|45934293|gb|AAS79232.1|	62	9.30e-18	KSIPPQCHCTDIRLNSCHSACKSCMCTRSMPGKCR
gi|154147103|emb|CAO82009.1|	62	9.30e-18	DSIPPICQCTDIRLNSCHSACKSCMCTRSMPGKCR
gi|169807808|emb|CAQ17032.1|	62	9.30e-18	DSIPPICQCTDIRLNSCHSACKSCMCTRSMPGTCR
gi|18541|emb|CAA48655.1|	64	2.90e-17	KSNPPQCRCSDMRLNSCHSACKSCICALSYPAQCF
gi|145586432|emb|CAM88858.1|	62	4.90e-17	DSIPPICQCTDIRLNSCHSACKTCMCTRSIPGTCR
gi|15529129|gb|AAK97768.1|	66	8.51e-17	KSIPPQCRCTDLRLDSCHSACKSCICTLSIPAQCV
gi|90762023|gb|ABD97865.1|	66	2.27e-16	KSIPPKCRCSDLRLNSCHSACKSCACTYSIPAQCY
gi|300396|gb|AAB26656.1|	39	5.62e-15	KSIPPQCRCSBLRLNSCHSECKGCICTFSIPAQCI
gi|300397|gb|AAB26657.1|	38	5.62e-15	KSIPPQCRCSBLRLNSCHSECKGCICTFSIPAQCI

The Motif 2 reveals that the query sequences CAL69282.1 has the pattern consist of "**DLRLNSCHSACKSCI"** amino acid in the respective order**.** In the pattern of AAK97767.1 amino acid **A** present in the tenth position of above mentioned pattern has been replaced with **E ("DLRLNSCHSECKSCI")** and the amino acid in the position tenth respectively. In the pattern of AAS79232.1, CAO82009.1 and CAQ17032.1 amino acid amino acids in the position of second, fifth, thirteenth and fifteenth respectinvely. in the above mentioned pattern has been replaced with **T, Q, L, N, S, F, S, S, E, I, V** and **A ("TDQLNSFSSEIVSAV** In the next pattern of ABK78689.1 all amino acids **P, N, N, P, K, A, C, P, R, N, C, D, T, R and I** (except the amino acid K present in the fifth position) present in the above mentioned pattern has been replaced with **V, L, K, L, R, G, D, K, E, E, K, F,**

ISSN: 0975-8585

K and V ("VLKLKRGDKEEKFKV ") amino acids respectinvely. In the next pattern of A32AAF18450.1 all amino **V, L, K, L, R, G, D, K, E, E, K, F, K and V** (except the amino acid K present in the fifth position all amino acids **N, P, I, C, I, N, C, C, S, G, Y, K, G , C** and **N** present in the above mentioned pattern has been replaced with **E, L, R, Q, V, D, F, L, T, K, A, N, E, V and I ("ELRQVDFLTKANEVI ")** amino acids respectinvely. In the next pattern of AAB24340.1 all amino acids **E, L, R, Q, V, D, F, L, T, K, A, N, E, V and I** present in the above mentioned pattern has been replaced with **G, A, K, G, G, R, C , I, W, E, G, T, N** and **V ("GAKGGRCIWGEGTNV ")** amino acids respectinvely(table 3)

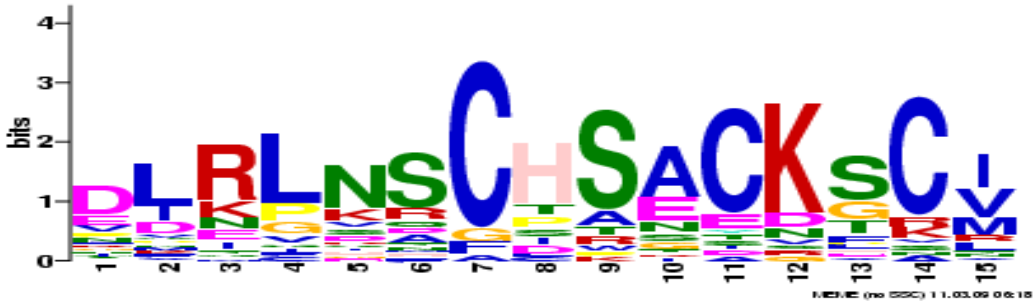

Figure 3: Sequence logo of motif 2

Motif 3 has the width of 11 sites and E-value of 1.5e-034. We have listed the Motif alignment in the following table with P-Value and the Start site of the Motif in the sequence (). Motif regions were coloured based on the biochemical properties of the amino acids. Most Positively charged amino acids (**K,R**) are indicated in red colour, Histidine , Glycine, Proline and Tyrosine are indicated in Pink, Orange, Brown and Turquoise colour respectively.

Motif 3 reveals that the query sequences AAK97770.1, AAK97765.1, AAK97768.1, ABD97865.1and CAA48655.1 has the pattern consist of "**ESSKPCCDQCA"** amino acid in the respective order**.** In the pattern of AAK97767.1 amino acid **P** (Hydrophobic amino acids) and **Q** present in the fifth and ninth position respectively of above mentioned pattern has been replaced with **A** and **H ("ESSKACCDHCA")** and the amino acid. 1 amino acid **D, R, H, I, P, K, K, Q, E** and **L** present in the first, second, third,fourth, fifth, senenth, eight, ninth, tenth and eleventh position respectively in the above mentioned pattern has been replaced with **S, D, D, N, V, C, N, G, C** and **L ("SDDNVCCNGCL")** and the amino acid in the position first, second, third, fourth, fifth, senenth, eight, ninth, tenth and eleventh respectinvely.

Table 4: MEME results of motif 3

NAME	START	P-VALUE	SITES
gi|15529133|gb|AAK97770.1|	42	2.85e-13	GHHQSTDEPSESSKPCCDHCACTKSIPPQCH
gi|15529123|gb|AAK97765.1|	42	2.85e-13	DHHQSTDEPSESSKPCCDQCACTKSIPPQCR
gi|15529129|gb|AAK97768.1|	43	2.85e-13	HHHESTDEPSESSKPCCDHCACTKSIPPQCR
gi|90762023|gb|ABD97865.1|	43	2.85e-13	RHHESTDEPSESSKPCCDQCACTKSIPPKCR
gi|18541|emb|CAA48655.1|	41	2.85e-13	KSDHQHSNDDESSKPCCDQCACTKSNPPQCR
gi|15529127|gb|AAK97767.1|	43	3.27e-12	HHHESTDEPSESSKACCDHCACTKSIPPQCR
gi|118026422|emb|CAL69282.1|	50	3.27e-12	HHHESTDEPSESSKACCDQCACTKSIPPQCR
gi|154147103|emb|CAO82009.1|	39	3.27e-12	NHHDSSDEPSESSEPCCDHCMCTDSIPPICQ
gi|145586432|emb|CAM88858.1|	39	1.56e-10	NHHDSSDEPSESSEPCCDLCMCTDSIPPICQ
gi|169807808|emb|CAQ17032.1|	39	1.56e-10	NHHDSSDEPSESSEPCCDLCMCADSIPPICQ
gi|45934293|gb|AAS79232.1|	39	1.61e-09	HHDDSSDEPSESSEPCCDSCICTKSIPPQCH

ISSN: 0975-8585

Figure 4 : Sequence logo of motif 3

When we compare the above mentioned motifs, all the three motifs are mainly consist of CYS residues. In between the CYS residues we found some amino acids such as **T, K, S, I, P, Q, R.** Generally Proteses especially Trypsin and Chymotypsin which is present in the mid gut of insect pests reacts with the PIs at particular site called as reactive site (Terra et.al, 1996). These reactive sites usually locked by disulfide bonds. Thus, we can conclude that the motifs which we have discovered are the reactive sites of PIs.

PHI BLAST:

The three motifs discovered from MEME tools were incorporated in to the PHI BLAST tool to identify the similar protein which have same motif like our query. The pattern **CT[KR]SIPP[QT]CRC** is closely related to Trypsin inhibtor of **Arabidopsis** thaliana (Mouse-ear cress), Arabidopsis lyrata subsp. petraea (Northern rock-cress) (Cardaminopsis petraea), Brassica napus (Rape), Clitoria ternatea (Butterfly pea), Descurainia sophia (Flixweed tansymustard) (Sisymbrium sophia), Sinapis alba (White mustard) (Brassica hirta) and Vitis vinifera (Grape) with **>50** of alignment score

MOTIF -02

The **DLRLNSCHS[AE]CKSC[IV]** pattern is closely related to Trypsin inhibtor of Arabidopsis thaliana (Mouse-ear cress), Arabidopsis lyrata subsp- petraea (Northern rock-cress) (Cardaminopsis petraea), Buthus occitanus mardochei (Moroccan scorpion), Brassica napus (Rape), Clitoria ternatea (Butterfly pea), Descurainia sophia (Flixweed tansymustard) (Sisymbrium sophia), Sinapis alba (White mustard) (Brassica hirta) and Vitis vinifera (Grape) with >50 of alignment score (Table 6).

Table 6: PHI BLAST results of motif 2

INHIBTOR	ORGANISM SOURCE	SIMILARITY %
Trypsin inhibitor 2	Sinapis alba (White mustard) (Brassica hirta)	100
Trypsin inhibitor 2	Sinapis alba (White mustard) (Brassica hirta)	100
Putative trypsin inhibitor 4	Arabidopsis thaliana (Mouse-ear cress)	87
Rapeseed putative trypsin inhibitor 1	Brassica napus (Rape)	91
Putative trypsin inhibitor 1	Arabidopsis thaliana (Mouse-ear cress)	96
Trypsin inhibitor ATTI-2	Arabidopsis thaliana (Mouse-ear cress)	96
Expressed protein (Putative tryspin inhibitor 4) (At2g43535) (Trypsin inhibitor 2)	Arabidopsis thaliana (Mouse-ear cress)	87

MOTIF-03

The **ESSK[PA]CCD[QH]CA** pattern is closely related to trypsin inhibtor of Arabidopsis thaliana (Mouse-ear cress), Arabidopsis lyrata subsp. petraea (Northern rock-cress) (Cardaminopsis petraea), Brassica napus (Rape), Buthus occitanus mardochei (Moroccan scorpion), Clitoria ternatea (Butterfly pea), Descurainia sophia (Flixweed tansymustard) (Sisymbrium sophia), Sinapis alba (White mustard) (Brassica hirta) and Vitis vinifera (Grape) with >50% of alignment score(table 7).

ISSN: 0975-8585

INHIBTOR	ORGANISM SOURCE	SIMILARITY %
Trypsin inhibitor 2	Sinapis alba (White mustard) (Brassica hirta).	100
Trypsin inhibitor 2	Sinapis alba (White mustard) (Brassica hirta).	100
Putative trypsin inhibitor 4	Arabidopsis thaliana (Mouse-ear cress).	87
Rapeseed putative trypsin inhibitor 1	Brassica napus (Rape).	91
Putative trypsin inhibitor 1	Arabidopsis thaliana (Mouse-ear cress).	96
Trypsin inhibitor ATTI-2	Arabidopsis thaliana (Mouse-ear cress).	96

From the above results we can conclude that most of the residues which we have taken as query are closed related to Typsin inhibitors. These inhibitors can block the activity of Trypsin of insect pests.

CONCLUSION

The role of PIs as defensive compounds against predators is particularly well established. The major role for serine PIs in animals is to block the activity of endogenous proteinases in tissues where this activity would be harmful. Inhibition of protease activity reduce the quantity of proteins that can be digested, thus the insects become weak with stunted growth and ultimately die. We have collected 30 PI sequences from NCBI and analysed the Motifs, Evolutionary relationship of the PI's using bioinformatics tools such as CLUSTALW and MEME. From this study we found that Motifs of PIs mainly comprises of CYS residues which have locked the reactive sites in it ("**CTRSIPGTCRC")**. In the above mentioned Motif of PI we can find the "ARG"in between the "CYS" residues (ARG is the reactive site of Trypsin inhibitors). Among the 30 sequences most of the sequences shares common motifs but some motifs are found with biochemically varied amino acids. This variation may due to the evolution of PIs.

REFERENCES

[1] Shulke RH, Murdock LL. Environ Entomol 1983; 12: 787-791.
[2] Chen MS, Johnson B, Wen L. Protein Express Purif 1992; 3: 41-49.
[3] Wolfson JL, Murdock LL. J Chem Ecol 1990; 16:1089-1102.
[4] Garciaolmedo SF. Plant. Oxford Survey Plant Mol Cell Biol 1987; 4: 275-334.
[5] Houseman JG, Downe AER, Philogene BJR. Can J Zool 1989; 67: 864-868.
[6] Lemos FJA, Xavier-Filno J, Campos FAP. Agri Biol Tecnoi 1987; 80: 46.
[7] Kimura M. Biosci Biotechnol Biochem 1995; 59: 2328-2329.
[8] León J. J Exp Bot 2001; 52: 1-9.
[9] Melville JC, Ryan CA. J Biol Chem 1973; 247: 415-3453.
[10] Yeh KW. Plant Cell Rep 1997; 16: 696-699.