

## In Silico Prediction of Three Dimensional Structure of Q699Z2, B5TGP9, D3JXJ8, F6K704 Carboxypeptidases from Some Lepidopteran Pests and analysis of Biopesticidal Activity of Kunitz-type Chymotrypsin Inhibitor Protein WCI-5 by Molecular Docking approach.

Anindya Sundar Panja<sup>1</sup>, Shakilur Rahman<sup>2</sup>,  
Bidyut Bandopadhyay<sup>2\*</sup> and Smarajit Maiti<sup>2</sup>

<sup>1</sup>Department of Biotechnology and Biochemistry.

Burdwan Institute of Management and Computer Science.

<sup>2</sup>Post Graduate Department of Biotechnology and Biochemistry. Oriental Institute of Science and Technology.

\*Corresponding author: Dr. Bidyut Bandopadhyay. E-Mail. [bidyut2006@gmail.com](mailto:bidyut2006@gmail.com)

[Received 12/11/2014, Accepted-08/12/2014]

### ABSTRACT

In this study, four full-length carboxypeptidase (CBP) proteins from some lepidopteran insects were sequenced (accession number, respectively). These sequences were further characterized in silico by subjecting them to multiple sequence alignment, phylogenetic tree construction, and protein motif analysis, revealing their identity to carboxypeptidase like proteins. Phylogenetic analysis of CBPs proteins along with other reported CBPs proteins divulged existence. Motif scan analysis of CBPs proteins revealed the presence of arginine-rich profiles in hr55, while lysine and alanine-rich profile was observed in hr53. Lysine, glycine and phenylalanine-rich profiles were found in hr56 protein. And alanine-rich profiles were noticed in hr51 protein. The three dimensional structures of CBP proteins were predicted by Swiss-modeller server based on multiple threading method. Validation of the protein was successfully done in ADIT server. The modelled structures were refined by energy minimization and their stereo chemical qualities were validated by PROCHECK and QMEAN server which confirmed the acceptability of the predicted models. Predicted CBP proteins were subjected to molecular docking by Hex 6.3 with Kunitz-type chymotrypsin inhibitor protein coded by WCI-5 gene and the high energy collision induced dissociation study confirmed the docking of the above proteins.

**Keywords:** Carboxypeptidase, Motif scan, Phylogenetic tree, Secondary structure, 3D structure, Ramachandran plot.

### INTRODUCTION

*Psophocarpus tetragonolobus* belongs to the family leguminosae which is a high yielding crop found in hot, humid equatorial countries like Philippines, Indonesia, Thailand, Sri lanka and India. *P.tetragonolobus* is also called as winged bean, goa bean and american bean. Various parts of *P.tetragonolobus* (leaves, flowers, roots) are edible and the beans are used as vegetables.

*P.tetragonolobus* have been proved that it has resistance against fungal and pest infection. The chymotrypsin inhibitor gene WCI-5 of *Phosphocarpus tetragonolobus* showed potential resistance against gut proteinase (Carboxypeptidase) of *Helicoverpa armigera* and *Helicoverpa zea*. The sequence of gut carboxypeptidase of *H. Armigera* is similar with

*Trichoplusia ni* (69%), *Loxostege sticticalis* (beet webworm) (54%), *Spodoptera frugiperda* (59%), *Mamestra configurata* (bertha armyworm)(80%), *Helicoverpa zea* (45%). A carboxypeptidase (EC number 3.4.16 - 3.4.18) is a protease enzyme that hydrolyzes (cleaves) a peptide bond at the carboxy-terminal (C-terminal) end of a protein or peptide. (Contrast with an aminopeptidase, which cleaves peptide bonds at the other end of the protein.) Humans, animals, and plants contain several types of carboxypeptidases that have diverse functions ranging from catabolism to protein maturation. However, most of the known carboxypeptidases are not involved in catabolism; they help to mature proteins (e.g., Post-translational modification) or regulate biological processes. In this classification system, carboxypeptidases that have a stronger preference for those amino acids containing aromatic or branched hydrocarbon chains are called carboxypeptidase A (A for aromatic/aliphatic). Carboxypeptidases that cleave positively charged amino acids (arginine, lysine) are called carboxypeptidase B (B for basic). Chymotrypsin Inhibitors (CI) are serine proteinase inhibitors of chymotrypsin. They are classified by numbers: CI-1, CI-2, and CI-3.

This bean has been called the "one species supermarket" because practically all of the plant parts are edible. The beans are used as a vegetable, but the other parts (leaves, flowers, and tuberous roots) are also edible. The tender pods, which are the most widely eaten part of the plant (and best eaten when under 1" in length), can be harvested within two to three months of planting. The flowers are often used to colour rice and pastries. The flavour of the beans has a similarity to asparagus. The young leaves can be picked and prepared as a leaf vegetable, similar to spinach. The roots can be used as a root vegetable, similar to the potato, and have a nutty flavour; they are also much richer in protein than potatoes. The dried seeds can be useful as a flour and also to make a coffee-like drink. Each of these parts of the winged bean provides a source of vitamin A, vitamin C, calcium, iron, and other vitamins [1]. The seeds contain 35% protein and 18% oil.

This paper reports in silico characterization of four full lengths CBP proteins from some pests by subjecting them to sequence alignment, Phylogenetic relationship analysis, motif analysis followed by three dimensional structure predictions and docking with kunitz-type chymotrypsin inhibitor. Two putative Kunitz-type chymotrypsin inhibitor genes (*WCI2* and *WCI5*) from winged bean (*Psophocarpus tetragonolobus* (L.) DC). *WCI5* and *WCI2* were expressed in *Pichia pastoris* and the recombinant proteins were assayed against various proteinases. Both the inhibitors strongly inhibited commercially available bovine chymotrypsin. More interestingly, gut proteinases of *Helicoverpa armigera* larvae that damage many important crop plants, were inhibited by *WCI2* and *WCI5*. Both of these proteinase inhibitors demonstrated significant reduction of growth of *H. armigera* larvae after feeding on inhibitor incorporated artificial diets. The inhibitory effects of *WCI2* and *WCI5* on activity of proteinases and larval growth make these proteins and their genes promising candidates for enhancing plant defence against *H. armigera* using transgenic plant [2]. The seeds of winged bean (*Psophocarpus tetragonolobus* L., DC.) that contain several proteinase inhibitors. Two-dimensional gel analysis of WB seed protein followed by activity visualization using a gel-X-ray film contact print technique revealed at least 14 trypsin inhibitors (TIs) in the range of 28-6 kD. Winged bean seed were individually assessed for their potential to inhibit the gut proteinases (HGP) of *Helicoverpa armigera*, a pest of several economically important crops, which produces at least six major and several minor trypsin/ chymotrypsin/ elastase-like serine proteinases in the gut. [3].

#### AIM AND OBJECTIVES

- To find out the sequentially similar sequences with gut carboxypeptidase of *H.armigera* and prediction of three dimensional structure of selected sequences.
- To study the pesticidal effect of *WCI-5* against the predicted structure of CBP proteins through molecular docking technique.

## MATERIALS AND METHODS

The amino acid sequences of Carboxypeptidase (CBP) of *Helicoverpa armigera* with Uniprot accession number 097389 is taken to run BLAST. And the amino acid sequences of five gut CBP proteins with assigned Uniprot accession number Q699Z2, B5TGP9, D3JXJ8, F6K704, Q3T905 of *Trichoplusia Ni*, *Loxostege sticticalis*, *Spodoptera frugiperda*, *Mamestra configurata*, *Helicoverpa zea* respectively, are taken to analysis phylogenicity as well as structure prediction.

### Three Dimensional Structure Prediction and Evaluation of 4 Gut Carboxy Peptidase Proteins

It was ascertained that the three dimensional structure (3D) of Gut carboxypeptidase proteins are not available in PDB database (<http://www.rcsb.org/>), hence an attempt has been made in the present study to determine the structure of four gut carboxy peptidase proteins from insects. The three dimensional structure of gut carboxy peptidase proteins has been predicted by iterative threading assembly refinement algorithm SWISS MODEL. The rough model generated was subjected to energy minimization using the steepest descent technique to eliminate bad contacts between protein atoms. Computations and implementations were carried out by the Swiss PDB Viewer version 4.0.3 (<http://expasy.org/spdv/>).

### Validations and Structural Motif Analysis

The backbone conformations of the predicted models were inspected by the Phi/Psi Ramachandran plot obtained from PROFUN server (<http://www.ebi.ac.uk/thornton-srv/databases/pdbfun/Generate.html>). The energy analysis was carried out by QMEAN server [23] (<http://swissmodel.expasy.org/qmean/cgi/inde.cgi>). The refined and validated models were submitted to ADIT database (<http://deposit.rcsb.org/adit/>). PDB files of Gut Carboxy Peptidase proteins were subjected to PDBsum server (<http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html>) for structural motif analysis.

### Docking Study

After obtaining the final model of CBP proteins, they are docked against the kunitz-type chymotrypsin inhibitor protein by the HEX 6.3 software.

## Free energy calculation

The free energy of binding is calculated using Swiss pdb-viewer approach. This approach is used to predict the free energy of binding for set of ligands to the receptor.

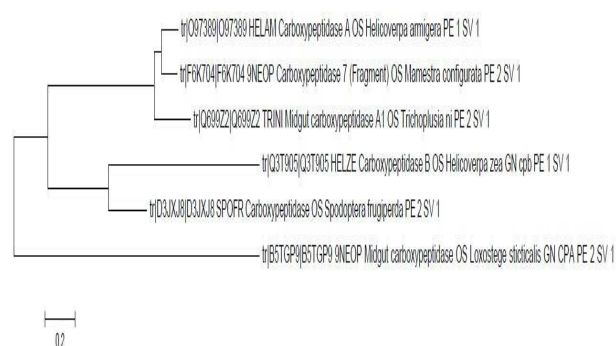
## RESULTS AND DISCUSSION

### Sequencing of CBP proteins from insects

The protein sequences of gut carboxy peptidases of *Helicoverpa armigera* accession number 097389 subjected to BLASTp search for reducing similarity with available sequences in uniprot database, which showed maximum similarity hit with gut carboxy peptidase like sequences of different species. These similar ortholog sequences of gut carboxy peptidase proteins were subjected to Phylogenetic tree constructed by never joining method by MEGA 6.3 software which showed maximum similarity with five gut carboxy peptidases. The gut carboxy peptidase of *Helicoverpa armigera* shares the common ancestor with these five gut carboxy peptidases.

### Phylogenetic relationship of CBP proteins

The protein sequences of CBP genes from those insects i.e., carboxypeptidases of *Helicoverpa zea*, *Loxostege sticticalis*, *Mamestra configurata*, *Spodoptera frugiperda*, *Trichoplusia ni* were subjected to BLASTp search for deducing similarity with available sequences in databases, which showed hits with *H. armigera* CBP protein. Phylogenetic tree was constructed based on aligned protein sequences using Neighbour-joining method. A tree was inferred by Bootstrap Phylogenetic inference by MEGA5.



**Fig. 1** Phylogenetic tree constructed based on carboxypeptidase protein sequences

Four sequences are taken for the prediction of secondary, tertiary structural classification by

Swiss modeller. Those were not present in PDB database. Molecular weight, isoelectric point, and amino acid percentage-rich region of four carboxypeptidase proteins were analyzed by PFAM tool. The deduced protein sequences of carboxypeptidase genes were subjected to motif scan search. Motif scan is integrated with PFAM. The chosen models of carboxypeptidases of *Loxostege sticticalis* (HR55), *Mamestra configurata* (HR53), *Spodoptera frugiperda* (HR56), *Trichoplusia ni* (HR51) from Swiss-modeler server were then subjected to energy minimization by Swiss-PdbViewer software for stabilizing their stereochemical properties. The energy value before and after energy minimization for each Carboxy peptidase proteins are shown in

**Table 2.**

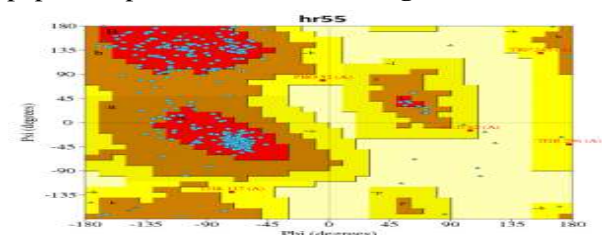
Protein Name	Energy KJ/Mole (before energy minimization)	Energy KJ/Mole (after energy minimization)
HR55	-16,617.049	-21,619.090
HR53	-9,361.223	-10,463.487
HR56	-14,159.820	-18,039.910
HR51	-16,409.434	-20,043.814

**Table 2:** Comparison of energy value of four CBP proteins before and after energy minimization

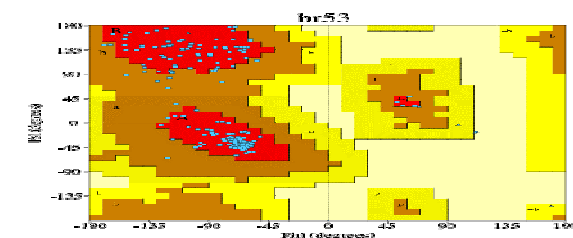
**Validation of the Predicted 3D Structure**

The stereochemical qualities of the predicted models of four carboxy peptidase proteins were validated by subjecting PDB files to PDB server and assessed by ADIT server. The Ramachandran plot statistics (% of the residues in the core region, allowed regions and in the disallowed region) for the four Carboxy peptidase models are shown in **Table 3** and **Fig. 2**. The shading on the plot represents different regions; the red areas correspond to the core regions representing the most favorable combinations of phi-psi values

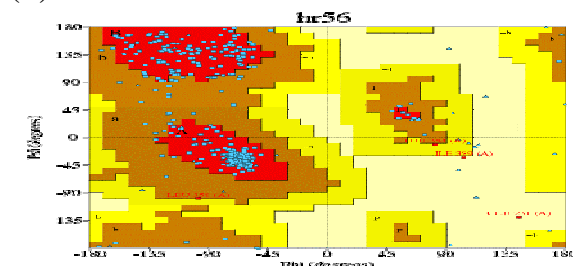
The Ramachandran plots of the modeled carboxy peptidase protein are shown in **Fig. 2**.



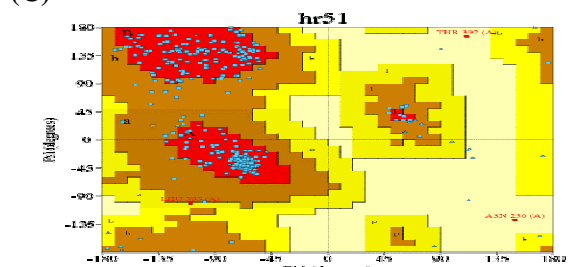
(A)



(B)



(C)



(D)

**Fig. 2:** Ramachandran plot of Gut Carboxy Peptidase proteins (A) hr55 (B) hr53 (C) hr56 (D) hr51.

The plot calculations on the 3D models of gut carboxy peptidase proteins were computed with the PROCHECK. Most favored regions are colored red, additional allowed, generously allowed, and disallowed regions are indicated as yellow, light yellow and white fields, respectively. 3D three dimensional (Color online)

Maximum likelihood of finding residues of protein (>80 %) in the core regions suggests better stereochemical quality. The results of the PROCHECK analysis indicate that a relatively low percentage of residues have phi/psi angles in the disallowed regions suggesting the acceptability of Ramachandran plots for carboxy peptidase proteins. The percentage of residues in the allowed/ core region were found to be 84.3, 88.9, 86.4, and 85.9 % for HR55, HR53, HR56 and HR51 respectively, while residues in disallowed regions were found to be 0.0, 0.0, 0.6, and 0.6 % for HR55, HR53, HR56 and HR51, respectively, as shown in Table 3. The stereo chemical quality of the predicted model was found to be satisfactory.

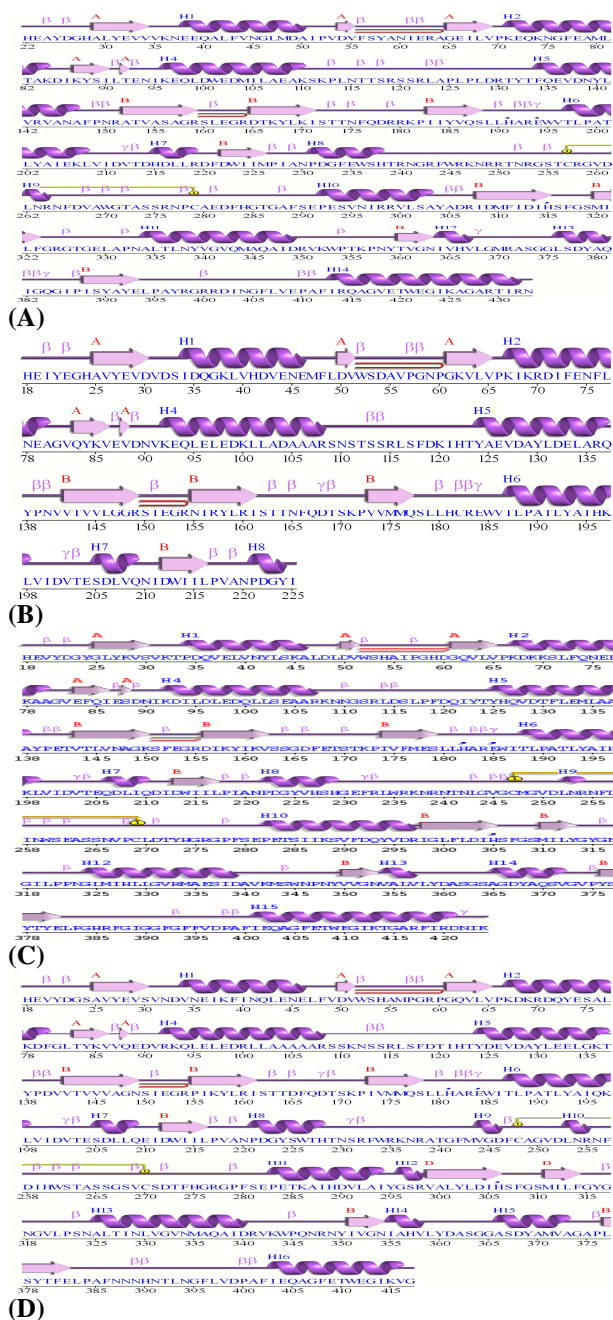


Fig. 4 Secondary structure of of (A) HR55 (B) HR53 (C) HR56 (D) HR51 protein of insects generated by PROfunc

**Table 3:-**The qualities of 3D structure of the four Carboxy peptidase proteins were further estimated by subjecting them in QMEAN Server(<http://swissmodel.expasy.org/qmean/cgi/ind ex.cgi>). The QMEAN Z score were found to be 0.61 (Z-score: -2.58), 0.73 (Z-score: -0.81), 0.62 (Z-score: -2.49), and 0.62 (Z-score: -2.53) for HR55, HR53, HR56 and HR51, respectively. The presence of significant QMEAN Z score for the four Carboxy peptidase proteins suggested the predicted model quality to be acceptable.

### 3D structure and motif analysis

The presence of  $\alpha$ -helix,  $\beta$ -sheets, turns, and coil were frequently observed in all the CBP proteins. CBP domain region mainly contain turns though sheets, coil, and helix were also observed. The overall predicted secondary structures of four CBP proteins are shown in Fig. 3.

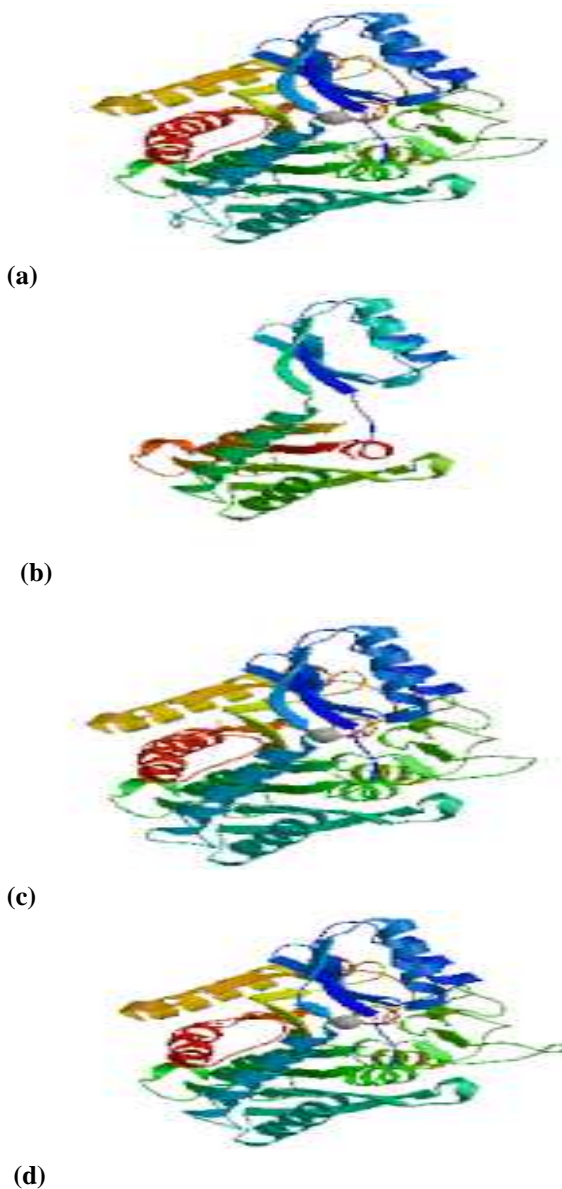


Fig. 3 The final 3D structure of a HR55 b HR53 c HR56 d HR51 protein of insects generated by SWISS modeler.

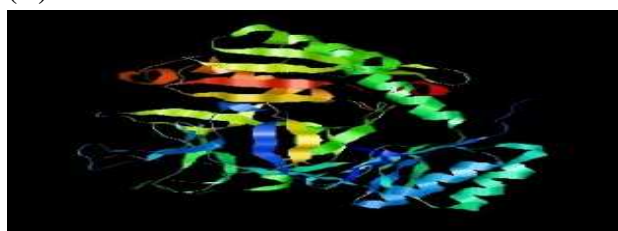
In case of hr55 protein arginine-rich (240-262 aa) profiles are found in motif scan output 9(Fig. 3a). Hr53 protein showed lysine and alanine-rich (95-107 aa) motif which consist a  $\alpha$ - helix (Fig. 3b). The hr56 protein contains lysine-rich profile (95–

135 aa) and glycine and phenylalanine-rich profile (383-415 aa) in motif scan analysis (Fig. 3c). The hr51 protein showed alanine-rich H4 helix (Fig. 3d).

Structural motifs	HR55	HR53	HR56	HR51
Sheets	2	2	2	2
Beta-alpha-beta unit	2	1	2	2
Beta hairpin	2	2	2	2
Beta bulges	5	3	5	5
Strands	12	8	12	12
Helices	14	8	15	16
Helix-helix interactions	12	5	10	11
Beta turns	43	21	38	37
Gamma turns	5	3	4	4

**Table 3** Structural motifs observed for CBP protein of insects

Results of PDBsum server revealed structural differences in four modeled CBP proteins. Structural motif beta–alpha–beta unit was present in all proteins in same number. Beta hairpins were present in all the CBP proteins in almost same number, while helix–helix interactions were observed in all CBP proteins and low in hr53. The observed numbers of different structural motif in four CBP protein models are shown in the docking was performed by the HEX in which protein and ligand were opened in docking software and was attached to the residue on the minimum distance position to the active site position. The ligand and target proteins were given as input to perform flexible docking. HEX 6.3 requires the receptor and ligand co-coordinator in PDB format to run a graphical user interface with relative speed compare to other complicated algorithm.



**Fig. 5** Docking of (A) hr55 (B) hr53 (C) hr56 (D) hr51 CBP proteins with chymotrypsin inhibitor.

Study on energy based on force field revealed that high energy induces dissociation of biomolecules whereas low energy exhibits structural and functional stability of natural biomolecules[4].

Pdb code	Before Docking	After Docking
(1JQG)	-21973.816	1352315904.0
(2C1C)	-31029.975	1885563628.0
(HR55)	-21,619.090	1397638.5
(HR53)	-10,463.487	1047884.063
(HR56)	-18,039.910	128050.984
(HR51)	-20,043.814	1305231590.0

**Table 4:-** Energy calculation of studied proteins before and after docking with WCI-5.

**FUTURE SCENARIO:** *P.tetragonolobus*, winged bean is edible as vegetable. It has natural antipesticial activity. By this study, more pests can be encountered through this bio-pesticide plant. It should be an aim to produce the transgenic plants by introducing WCI-5 pesticidal peptide gene to encounter those pests which affects agriculture as well as animals. As six gut carboxypeptidases of those are successfully docked with kunitz-type chymotrypsin gene, we can expect that researchers will find out better ways to fight against these pests.

## CONCLUSION:

Four CBP proteins namely hr55, hf53, hr56 and hr51 were predicted. In silico analysis were carried out by sequence alignment, phylogenetic tree construction and motif analysis. In the phylogenetic tree these CBP proteins were found to be close to *Helicoverpa armigera*. This paper reports for the first time the in silico prediction of three dimensional structure of CBP proteins. The 3D structures were generated by assembly simulations from online available Swiss-Pdb modeller server. The selected models were refined by energy minimization by Swiss-Pdb Viewer. The stereochemical qualities of predicted models were validated by PROCHECK server and Ramachandran plot analysis, which suggested that predicted models are satisfactory. Through gene ontology (GO) study, predicted proteins show their carboxypeptidase, protease and hydrolase activity. Predicted protein models were subjected to docking with kunitz-type chymotrypsin inhibitor (coded by WCI-5 gene). After analysing the docking figures we can easily conclude that chymotrypsin inhibitor can encounter these carboxypeptidase proteins. The results from energy based on force field (table 4) using Swiss pdb-viewer showed that all the selected proteins have exhibited high energy level after docking with WCI-5. So, it can be concluded that CBP-WCI-5 docked proteins must be degraded and thus CBP lost their peptidase activity. As predicted CBP protein models have emerging similarity with the carboxypeptidase of *Helicoverpa armigera* and inhibitor activity of winged bean to this pest has been proven before, we can hope that it will be helpful for researchers to discover pesticides to encounter these threats: *Trichoplusia Ni*, *Loxostege sticticalis*, *Spodoptera frugiperda*, *Mamestra configurata*, *Helicoverpa zea*.

## REFERENCE

1. <http://www.healthaliciousness.com/nutritionfacts/nutrition-comparison.php> (for winged bean seeds, leaves, and tubers.).
2. Telang, Manasi A. Giri, Ashok P.; Pyati, Prashant S.; Gupta, Vidya S.; Tegeger, Mechthild; Franceschi, Vincent R. February 15, 2009. Winged bean chymotrypsin inhibitors retard growth of

*Helicoverpa armigera*. *Gene* vol. 431 issue 1-2. p. 80-85.

3. Giri, Ashok P.; Harsulkar, Abhay M.; Ku, Maurice S.B.; Gupta, Vidya S.; Deshpande, Vasanti V.; Ranjekar, Prabhakar K.; Franceschi, Vincent R. July, 2003. Identification of potent inhibitors of *Helicoverpa armigera* gut proteinases from winged bean seeds. *Phytochemistry* vol. 63 issue.5. p. 523-532.
4. Pittenauer E, Allmaier G. feb-2009, High-energy collision induced dissociation of biomolecules: MALDI-TOF/RTOF mass spectrometry in comparison to tandem sector mass spectrometry. *Pubmed*, 12(2):137-55.