Journal of Food Legumes 24(2): 138-141, 2011

In-silico analysis and homology modelling of coat-protein of Mungbean Yellow Mosaic India Virus

NAVNEET KUMAR YADAV[1], SARIKA[2], M.A. IQUEBAL[2] and MOHD. AKRAM[2]

[1]Institute of Bioscience and Biotechnology, Chhatrapati Shahu Ji Maharaj University, Kanpur – 208 024, India; [2]Indian Institute of Pulses Research, Kanpur-208 024, Uttar Pradesh, India;E-mail: navneet.bioinfor @gmail.com

(Received: February 9, 2011; Accepted: April 23, 2011)

ABSTRACT

Mungbean is affected by several diseases incited by viruses, fungi, bacteria and nematodes. In this study, coat-protein sequence was analyzed and modelled to explore properties and structure of *Mungbean yellow mosaic India virus* (MYMIV). Since no structural information is available for majority of protein sequences available in Protein Data Bank, computational methods for protein structure prediction have been of much interest in recent years. Physico-chemical properties of the protein under study were computed. The grand average of hydropathy (GRAVY) value and instability index were found to be -0.646 and 45.75, respectively, supporting the protein to have better interaction with water. Comparative homology modelling was performed with ModWeb, ESyPred and Swiss Model. It was observed that the model generated by ModWeb was most acceptable with 90.2% of the residues in most favoured region. The model was validated using Structural Analysis and Verification Server (SAVS). Coat-protein structure of MYMIV obtained through computational modelling was found congruent with their protein structure obtained by X-ray crystallography or NMR.

Keywords: ESyPred, Homology modelling, ModWeb, Mungbean yellow mosaic India virus, Ramachandran plot, Swiss Model

Mungbean or green gram, scientifically known as *Vigna radiata* L Wilczek, is one of the important short duration pulse crops of Indian origin. It is cultivated across the country throughout the year with an area and production of 3.77 mha and 1.52 mt, respectively (AICRP on MULLaRP 2009). Its seeds contain approximately 25-28% protein, 1-1.5% oil, 3.5-4.5% ash and 62-65% carbohydrates on dry weight basis. Lysine content is high in mungbean, making its protein an excellent complement to rice in terms of balanced human nutrition.

Mungbean is known to be affected by several diseases incited by viruses, fungi, bacteria and nematodes. Of the various viral diseases, yellow mosaic disease is widely distributed and most destructive. The disease has been reported from various countries. Yellow mosaic disease of many legumes in India and other South Asian countries is caused by whitefly (*Bemisia tabaci* Genn.). The vector transmits geminiviruses belonging to the family *Geminiviridae* and genus *Begomovirus*. Mainly two viruses, *viz.*, *Mungbean yellow mosaic virus* (MYMV) and *Mungbean yellow mosaic India virus* (MYMIV), have been reported to be the causal agents of yellow mosaic disease in mungbean in India (Pant *et al*. 2001, Malathi and John 2008).

Experimental determination of protein structure through X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy is time consuming and very costly. Protein Data Bank is a repository for three-dimensional structural data of large biological molecules submitted by biologists and biochemists from around the world. Still, majority of protein sequences have no structural information as the number of unique structural folds that proteins adopt is limited and number of experimentally determined new structures is increasing exponentially. Therefore, it is necessary to bridge this 'structure knowledge gap'. Computational methods for protein structure prediction have received attention in recent years. Various approaches have been followed in this context. The three-dimensional structures of proteins in a family are obvious to be conserved more than their sequences. Hence, if one detects the similarity between two proteins at the sequence level, structural similarity can usually be assumed further. Moreover, it may happen that proteins that have no detectable sequence similarity may have similar structures. In case of animal and human beings, many proteins structure-function relationships can be deduced from a reasonable model, which may further be used for successful drug design. Thus the present study was conducted to predict the structure of coat protein of MYMIV, which may help in understanding the protein function, antigenic determinants for antibody production and in taxonomic studies.

MATERIALS AND METHODS

Fasta format for sequence of coat-protein of MYMIV with GenBank accession number ACV04859 was traced from National Centre for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov). Fasta format of template sequences with at least 30% sequence identity were downloaded from Research Collaborative for Structural Bioinformatics-Protein Data Bank (RCSB-PDB) after performing Basic Local Alignment Search Tool Protein (BLASTP). Template sequences in .pdb extension were retrieved from RCSB. Deep-View (Swiss-PdbViewer), the bioinformatics tool for modelling

of coat-protein of MYMIV was used in the present study. ESyPred3D, a new automated homology modelling programme as well as ModWeb was also employed to compare the results of the analysis.

The general scheme adopted for homology modelling is shown in Fig 1.

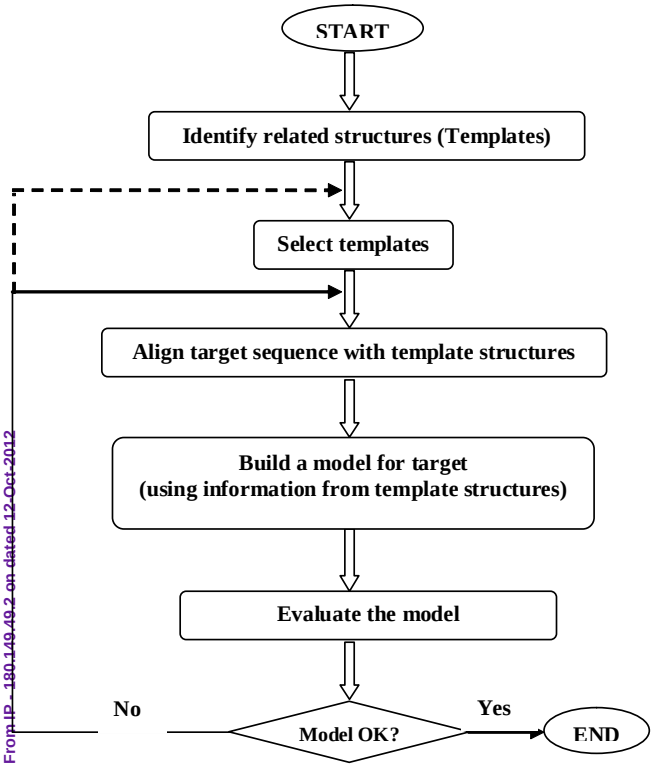

Fig 1. Diagrammatic representation of homology modelling

RESULTS AND DISCUSSION

***Physico-chemical characterization*:** To compute the physico-chemical characterization of sequence with 257 bp length under study, Expasy's ProtParam server (Gasteigers 2005) was used, which allows the computation of various physical and chemical parameters for a given protein. Molecular weight (MW), theoretical isoelectric point (pI), total number of positively (+R) and negatively charged residues (-R), extinction coefficient (EC) (Gill and Von Hippel 1989), optical density (OD) instability index (II) (Guruprasad *et al*. 1990), aliphatic index (AI) (Ikai 1980) and grand average hydropathy (GRAVY) (Kyte and Doolittle 1982) are represented in Table1.

Table 1. Physicochemical characterization of coat-protein

Molecular weight (Da)	pI	-R	+R	EC	OD	II	AI	GRAVY
29993.4	9.94	20	42	37360	1.25	45.75	61.44	-0.646

The extinction coefficient indicates how much light a protein absorbs at a certain wavelength. The instability index provides an estimate of the stability of protein in a test tube. The coat protein under study was found to be quite unstable. The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains. GRAVY value, which is calculated as the sum of hydropathy values of all the amino acids divided by the number of residues in the sequence, for the protein under study was -0.646, indicating the possibility of better interaction with water.

Disulphide bonds play an important role in determining the functional linkages. CYS_REC program for predicting SS-bonding states of cysteines and disulphide bridges in protein sequences has been used in our study. CYS_REC identified the position of cysteins at 69, 73, 92, 111, 141, 194 and 234. The most probable pattern of pairs found was 69-73 and 111-194.

Hydropathy plot of protein sequence was drawn by Winpep (Hennig 1999) and results are shown in Fig 2.

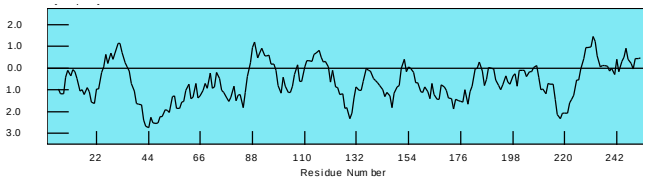

Fig 2. Hydropathy plot of coat-protein

Secondary structure prediction: SOPMA (Geourjon and Deleage 1995) was employed for calculating the secondary structural features of protein sequences in the present study. The results are presented in Table 2.

Table 2. SOPMA results of the coat-protein

Parameters	% content	Parameters	% content
Alpha helix	18.29%	Beta turn	6.23%
3_{10} helix	0.00%	Bend region	0.00%
Pi helix	0.00%	Random coil	50.58%
Beta bridge	0.00%	Ambiguous states	0.00%
Extended strand	24.90%	Other states	0.00%

Modelling of the sequence: ESyPred3D (Lambert *et al*. 2002), Swissmodel (Arnold *et al*. 2006) and ModWeb (Sali and Blundell 1993) were employed to perform modelling of the coat-protein under study. ESyPred3D is an automated homology modelling programme with the increased alignment performances of a new alignment strategy using neural networks. Swiss-Model can be accessed via ExPASy web server, or from the DeepView (Swiss Pdb-Viewer) programme. ModWeb is a web server for automated comparative protein structure modelling which accepts one or more sequences to calculate models for them based on the best available template structures from the Protein Data Bank (PDB). The comparative

results from these three servers are presented in Table 3. Swiss model and ESyPred3D showed 90.3 and 89.9% residues, respectively in the most favoured region. The result from ModWeb was found to be the best for the residues in which generously allowed and disallowed regions were 0.0%.

Table 3. Comparative analysis of the models obtained from ESyPred3D, ModWeb and Swiss-model program.

Server	Details of Residues	% content
ESyPred3D	Residues in the most favored Region	89.9%
	Residues in additionally allowed region	8.5%
	Residues in generously allowed region	0.5%
	Residues in disallowed region	1.0%
ModWeb	Residues in the most favored Region	90.2%
	Residues in additionally allowed region	9.8%
	Residues in generously allowed region	0.0%
	Residues in disallowed region	0.0%
Swiss model	Residues in the most favored Region	90.3%
	Residues in additionally allowed region	9.0%
	Residues in generously allowed region	0.7%
	Residues in disallowed region	0.0%

***Validation of the model*:** Evaluation of model quality is a crucial step in homology modelling. Once the homology modelling procedure is over, the final model needs to be inspected using validation tools in order to confirm whether the model's stereochemistry is reasonably consistent with typical values found in crystal structures. Persistent problems may suggest a problem with the alignment used to build the model; manual adjustments to the alignment may be necessary, particularly in the loop areas, followed by a rebuilding of the model. Structural Analysis and Verification Server (SAVS) were used for evaluation of model quality.

The Ramachandran plot (Fig. 3) shows the phi-psi torsion angles for all residues in the structure (except those at the chain termination). Glycine residues are separately identified by triangles as these are not restricted to the regions of the plot appropriate to the other side chain types. The darkest areas correspond to the "core" regions representing the most favorable combinations of phi-psi values. Ideally, one would hope to have over 90% of the residues in these "core" regions. The percentage of residues in the "core" regions is one of the best guides to stereo-chemical quality. Ramachandran plot in PROCHECK validation package was used to assess the quality of the modelled structure. 90.3% of the residues were in the core region. The model found was the best one having maximum core region and less disallowed region with minimum energy.

The average Z score, Z score RMS and distribution of atomic Z scores is represented in Fig 4. The RMS Z-score of the protein under consideration as found in PROVE of SAVS

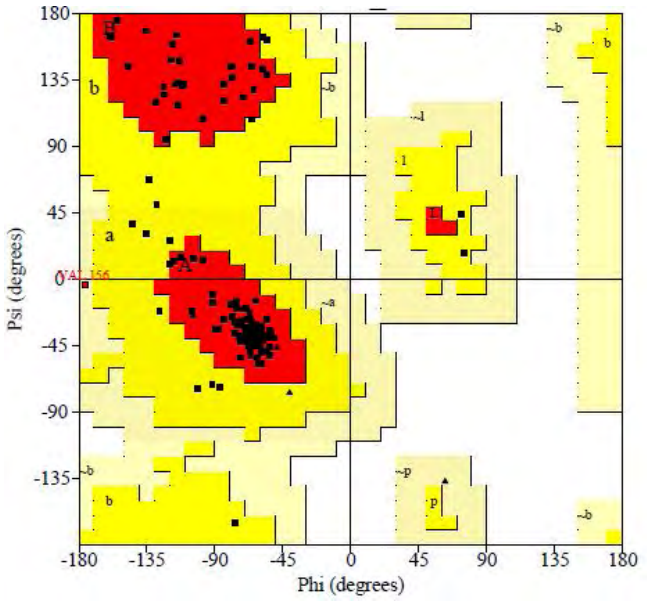

Fig 3. Ramachandran plot in PROCHECK

server was 1.716, indicating good model quality. The predicted structure conformed well to the stereochemistry, indicating reasonably good quality.

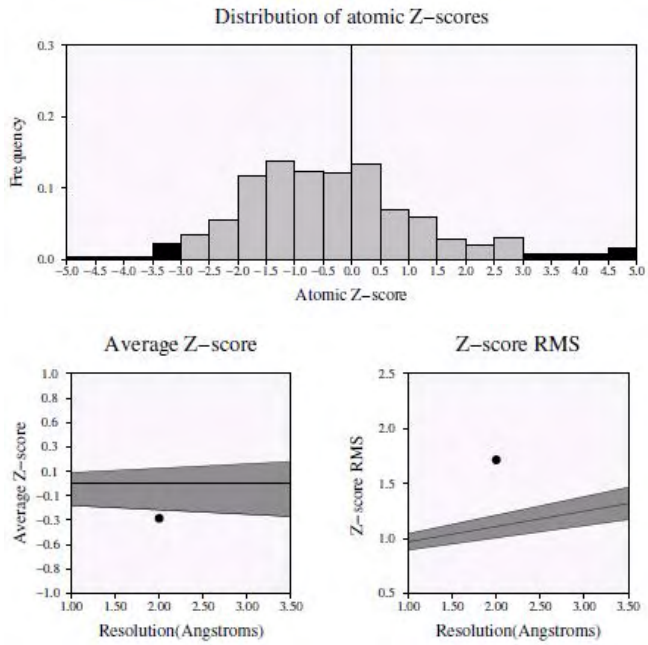

Fig 4. Distribution of atomic Z scores, average Z score and Z score RMS

A comparison of the results obtained showed that the model generated by ModWeb was more acceptable in comparison to ESyPred3D and Swiss Model. The final modelled structure for the coat protein under study visualized by Chimera software is illustrated (Fig. 5).

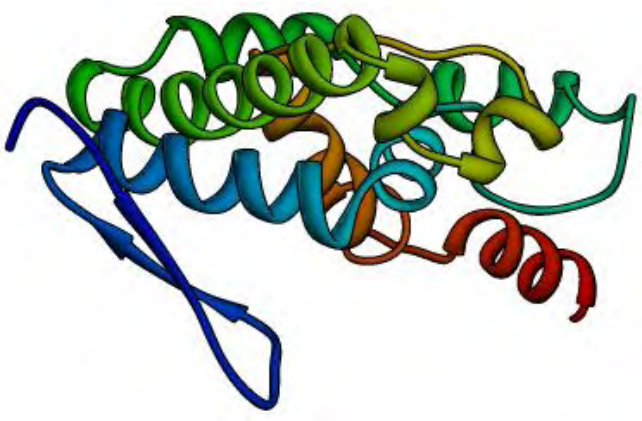

Fig 5. Ribbon modelled structure of coat protein of MYMIV

In this study, physico-chemical characterization was performed by computing theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average hydropathy (GRAVY) of coat-protein of MYMIV. The values of GRAVY and instability index were found to be -0.646 and 45.75, respectively. The protein seemed to interact well with water. Further, comparative homology modelling was performed by three different servers, *viz*. ModWeb, ESyPred and Swiss Model with the observation that the model generated by ModWeb was the most accurate, of high quality and acceptable with 90.2% of the residues in most favoured region. The model was validated using protein structure checking tools of SAVS. The homology modelling may provide a good foundation for functional analysis. Coat-protein structure of MYMIV obtained through computational modelling in this study was as good as their protein structure that could be obtained by X-ray crystallography or NMR. The study may help in understanding the protein function, number and types of epitopes, immunogenic portions, and suitability for antibody production, taxonomic studies, evaluation studies and virus diagnostics.

REFERENCES

Annual Report. 2009. AICRP on MULLaRP, IIPR, Kanpur 208 024

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

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD and Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker (Ed), The Proteomics Protocols Handbook, Humana Press. Pp 571-607.

Geourjon C and Deléage G. 1995. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Computer Applications in the Biosciences **11:** 681-684.

Gill SC and Von Hippel PH. 1989. Extinction coefficient. Analytical Biochemistry **182:** 319-328.

Guruprasad K, Reddy BVP and 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 Engineering **4:** 155-164.

Hennig L. 1999. WinGene/WinPep: User-friendly software for the analysis of aminoacid sequences. Biotechniques **26:** 1170-1172.

Ikai AJ. 1980. Thermo stability and aliphatic index of globular proteins. Journal of Biochemistry **88:** 1895-1898.

Kyte J and Doolottle RF. 1982. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology **157:** 105- 132.

Lambert C, Leonard N, De Bolle X and Depiereux E. 2002. ESyPred3D: Prediction of proteins 3D structures. Bioinformatics **18:**1250-1256.

Malathi VG and John P. 2008. Characterization, diagnosis & management of plant viruses. In: GP Rao, PL Kumar and RJ Holguin-Pena (Eds), Vegetables and Pulse crops (Volume 3). Stadium Press LLC, Texas, USA. Pp 97-123.

Pant V, Gupta D, Choudhury NR, Malathi VG, Varma A and Mukherjee SK. 2001. Molecular characterization of the Rep protein of the blackgram isolate of Indian mungbean yellow mosaic virus. Journal of General Virology **82:** 2559–2567.

Sali A and Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology **234:** 779-815.