Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env. Pharmacol. Life Sci., Vol 7 [8] July 2018 : 26-35

Online ISSN 2277-1808
Journal's URL:http://www.bepls.com
CODEN: BEPLAD
Global Impact Factor 0.876
Universal Impact Factor 0.9804
NAAS Rating 4.95

ORIGINAL ARTICLE **OPEN ACCESS**

In-silico Analysis & Homology modelling of Papaya Ring Spot Virus & Papaya Leaf Curl virus coat protein

Jay Patel[1] and Rishee K. Kalaria[2]*
[1].B.Sc. student, Aspee Shakilam Biotechnology Institutes, Navsari Agricultural university, Athwa Farm, Surat-39500
[2].Aspee Shakilam Biotechnology Institutes, Navsari Agricultural university, Athwa Farm, Surat-39500
***Corresponding Author:** risheekal@nau.in

ABSTRACT
The papaya crop is severely affected by PRSV, PapMV & PLCV worldwide. In this study, coat protein sequences of PRSV-P, PRSV-W, PapMV & PLCV were analyzed and modelled to explore properties and structure. Comparative homology modelling & Ab initio were performed by different servers, viz. ModWeb, Phyre2,Swiss Model & I-TASSER with the observation that the model generated by modweb for both PRSV-P & PRSV-W has most accurate, of high quality and acceptable highest percent residue in most favoured region 91% and 81% respectively. While PLCV, overall quality was 1.5 in I-TASSER with 52% percent residue in most favoured region. The values of GRAVY were -0.590, -0.124, -0.933 & -0.943 PLCV, PapMV, PRSV-P & PRSV-W respectively indicating the possibility of better interaction with water. The models were validated using protein structure checking tools PROCESS. These structures will provide a good foundation for functional analysis of experimentally derived crystal structures.
Keyword: PRSV, PLCV, ModWeb, I-TASSER & *PROCESS*

Received 11.05.2018 Revised 10.06.2018 Accepted 30.06.2018

INTRODUCTION

Papaya (*Carica papaya* L) belongs to the family *Caricaceae* and one of the most significant fruit crops grown widely in tropical and subtropical lowland regions ie India, Indonesia, China etc [1]. In India, among different fruit crops grown, papaya cultivation fifth position in terms of area and production [2]. It is a richest source of vitamin A, B, C and proteolytic enzymes viz. papain, chymopapain etc. Beta carotene helps in avoiding of cancer, diabetes, and heart diseases. Ripend fruits are frequently eaten fresh and can be processed into jam, jelly and candy. It is also used in the pharmaceutical and cosmetic industries [3].

Presently, the major constraint in the cultivation and production of papaya is its susceptibility to a number of diseases and particularly the disease caused by mainly serious viral disease including papaya ring spot disease, Papaya mosaic diseases, Papaya leaf curl diseases [4]. The viruses most often reported in this crop are *Papaya ring spot virus* (PRSV), a potyvirus *Papaya mosaic virus* (PapMV), a potexvirus, and Papaya leaf curl virus, a potyvirus (PLCV). PRSV, PapMV & PLCV infection are reported to occur in each area of the country where papaya is grown irrespective of the agroclimatic conditions and these diseases can result in crop losses up to 85-90% [5]. PRSV, a member of the genus *Potyvirus*, is further classified into two types: type P (PRSV-P), which infects cucurbits and papaya, and type W (PRSV-W), which infects cucurbits but not papaya. The biotypes P and W are serologically indistinguishable [1]. Viruses are transmitted by several species of aphids in a nonpersistent manner to a diverse host range. The common symptoms of viruses are prominent mosaic and chlorosis of leaf lamina, wet-oily streaks on the petioles and on the tree trunk, and complete distortion of young leaves. The fruit exhibits bumps and the classic "ringspot". This leads to 50 % or even more decline in fruit production [3].

The natural spread of PRSV, PapMV & PLCV are quick; therefore, the viruses may infect up to 100% of plants in a field. The disease is so devastating that farmers have stopped growing papaya in harshly affected areas. Infected plants lose vigor and become stunted. Production of fruit progressively later stages is severely reduced and of poor quality, owing to the presence of ringspots and generally lower

sugar concentrations[3]. Such impact of leaf curl disease in papaya has also led to overall reduced fruit production.

Characterization of PapMV coat protein structure by X-ray crystallography and cryo-EM 3D reconstruction [6]. Experimental determination of PRSV & PLCV coat 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 across the world. Still, majority of protein sequences have no structural information as the number of unique structural folds that proteins adopt is partial and number of experimentally determined new structures is increasing exponentially. Therefore, it is essential to link this 'structure knowledge gap'. Computational methods for protein structure prediction have received attention in recent years. Thus the present study was conducted to predict coat protein structure of PRSV & PLCV and PapMV as reference, which may help in understanding the protein structure function, physiochemical diversity among them.

MATERIAL AND METHODS

Sequence retrieval and Structure prediction:

Coat-protein(CP) fasta format sequences of PRSV-P, PRSV-W, PapMV & PLCV with GenBank accession number AEC04846.1, AOZ60527.1, NP_044334.1 & CUH74568.1 were traced from NCBI(www.ncbi.nlm.nih.gov). Fasta format of template sequences with at least 30% sequence identity were downloaded from RCSB-PDB after performing BLASTP. Template sequences in .pdb extension were retrieved from RCSB. PHYRE2 is an automated homology modelling programme with the increased alignment performances of a new alignment strategy using rely on profiles or hidden Markov models (HMMs). 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).

For homology modelling, PHYRE2, MODWEB and SWISS MODEL web servers were used for analysis while for Ab initio modelling, I-TASSER an online server, which is a hierarchical method for protein structure prediction.

Physico-chemical characterization:

The physicochemical characterization of sequences were carried out by Expasy's ProtParam server and SOPMA server [7,8]

Structure validation:

Evaluation of built model quality were analyze through amino acid region in Ramachandran plot in PROSESS web server [7,8].

RESULT AND DISCUSSION

BLASTP & Modelling of the sequence:

After performed BLAST-P for CP of PRSV-P, PRSV-W, PapMV & PLCV, It was observed that both CP of PRSV-P & PRSV-W showed 57% identities with Chain A, Structure Of Watermelon Mosaic Virus Potyvirus (5DOV_A).These both sequence were further carried forward for homology modelling by ModWeb,Swiss model & Phyre2 while PLCV showed 22% identities with Chain A, Vamp7 Longin Domain Hrb Peptide Complex(2VX8_A).CP of PLCV were further carried forward for Ab initio modelling by I-TASSER. PapMV had recently uploaded the PDB structure Chain A, Crystal Structure Of Papaya Mosaic Virus Capsid Protein (4DOX_A & 4DOX_B).For homology modelling, PHYRE2, MODWEB and SWISS MODEL were employed to perform modelling of PRSV-P & PRSV-W coat-protein under study. Prajapati and Bhagat, 2012 also performed BLAST-P for Clostridium botulinum protein sequence[8]. The comparative results from these three servers are presented in Table 1& Fig:2.

PHYRE2 and Swiss Model built model that showed 100% confident for 64% & 62.61% sequence identity residues for PRSV-P while 63% & 57.66% sequence identity residues for PRSV-W respectively in the most favoured amino acid region. The result from ModWeb built model that showed 23% for PRSV-P and 28% PRSV-W(Table 1& Fig:2). The result from PHYRE2 and Swiss Model were found to be the best for the residues coverage in PRSV-P & PRSV-W**.** Yadav et.al. 2011 also predict homology model using Modweb of MYMIV CP . The result from ModWeb was found to be the best for the residues in which generously allowed and disallowed regions were 0.0%.

I-TASSER is a hierarchical approach to protein structure and function prediction. It first identifies structural templates from the PDB by multiple threading approach, with full-length atomic models constructed by iterative template fragment assembly simulations. I-TASSER built model for PLCV that showed model-1 with highly confident score as best C-sore -2.68(Table:2, Fig:3).

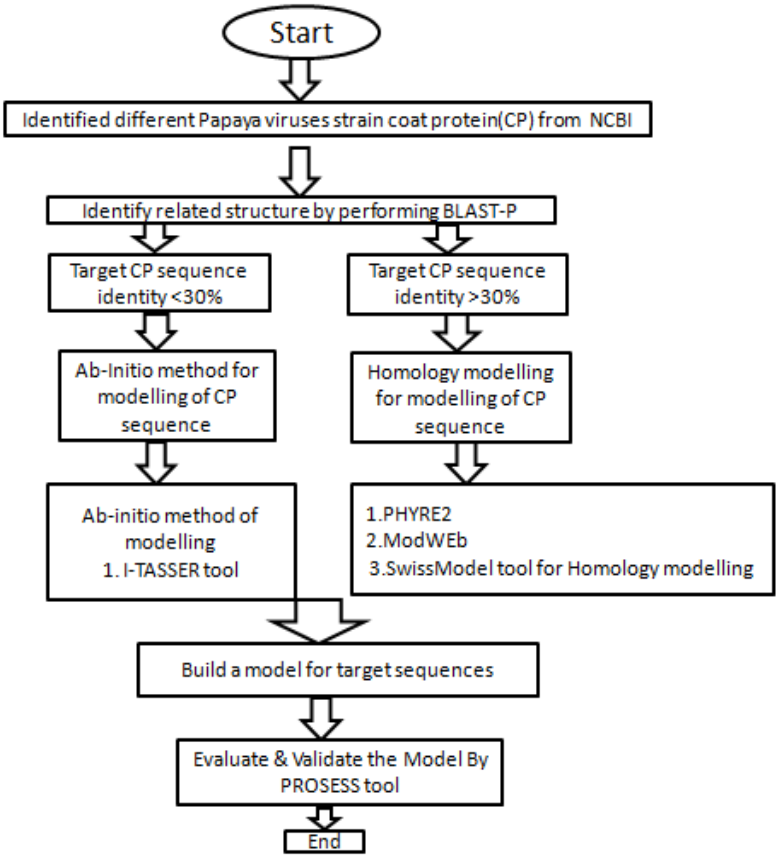


Fig: 1 Diagrammatic representation of homology & Ab initio modelling for different coat protein of Papaya viruses

Physico-chemical characterization:

To compute the physicochemical characterization of CP of PRSV-P, PRSV-W, PapMV & PLCV under study, Expasy's ProtParam server [9] was used, which allows the computation of various physical and chemical parameters for a given protein. Grand average hydropathy (GRAVY) values of all the amino acids divided by the number of residues in the sequence, for the protein under study was -0.590, -0.124, -0.933 & -0.943 PLCV, PapMV, PRSV-P & PRSV-W respectively indicating the possibility of better interaction with water (Table 3). Yadav et.al. 2011 also compute the physicochemical characterization of MYMIV CP sequence[7]. Prajapati and Bhagat, 2012 also compute for protein-ligand interactions against Clostridium botulinum[8]. Similarly Sahay & Shakya , 2010 studied physicochemical characterization of different antioxidant protein in spinach [10].

Secondary Structure Prediction:

SOPMA was employed for calculating the secondary structural features of given protein sequences in the present study. Yadav et.al. 2011 also used SOPMA for MYMIV CP sequence. The results revealed that Random coil dominated among secondary structure elements followed by Alpha helix, Extended strand and beta turns for all sequence (Table:4). Similarly Prajapati and Bhagat, 2012 observed that alpha helix dominated among secondary structure elements followed by extended strand, random coil and beta turns in their sequences. Similarly Sahay & Shakya , 2010 also studied secondary structure of spinach antioxidant proteins[10].

***Validation of the model*:**

Evaluation of model quality is a critical step in homology or Ab initio modelling. Once the model built, the final model require to be inspected using validation tools in command to confirm whether the model's stereochemistry is reasonably reliable with typical values originate in crystal structures.

Ramachandran plot calculation in PROCESS validation package was used to measure the quality of the modelled structure provided by Phyre2, Modweb, Swiss model and I-TASSER[11] calculations. The Ramachandran plot 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.

For PRSV-P & PRSV-W, over all quality were highest and same in ModWeb & Swiss model 4.5 respectively while highest percent 91% and 88% respectively) residue in most favoured region covered in modweb for both PRSV-P & PRSV-W(Table:5, Fig 4). The model found was the best one having maximum core region and less disallowed region with minimum energy While PLCV, over all quality was 1.5 in I-TASSER with 52% percent residue in most favoured region (Table:6, Fig 5). The model found was the poor one having minimum core region .Yadav et.al. 2011 predict MYMIV model that was the most accurate,of high quality and acceptable with 90.2% of the residues in most favoured region[7]. Similarly Sahay & Shakya , 2010 also studied also evaluate the predicted protein model by procheck using Ramachandran plot[10]. Gava *et al*. [12] also built model for capsid protein monomer of the PCV2 by I-TASSER online server for predicting region for antibodies recognition site.

Table: 1(A) Homology modelling of PRSV-P & PRSV-W CP using PHYRE2

PRSV-P			PRSV-W		
Template	**confidence**	**Percent Identity**	**Template**	**confidence**	**Sequence Identity**
C5odvB_	100	64%	C5odvB_	100	63%
C5fnlA_	97.6	16%	C5fnlA_	97.4	17%
C5a2tg_	97.4	16%	C5a2tg_	97.2	17%
C4doxA_	96.3	17%	C4doxA_	96.1	18%

Table :1 (B) Homology modelling of PRSV-P & PRSV-W CP using ModWeb

Sr. No	Parameters	PRSV-P	PRSV-W
1	Target region	68-171	12-150
2	Protein length	289	287
3	Template PDb code	4rm4A	3hy5A
4	Template region	110-232	50-172
5	Sequence identity	23.00%	28.0%
6	E-value	0.22	0.78
7	GA341	0.13	0.11
8	MPQS	0.369762	0.587921
9	z-DOPE	-0.03	-0.2
10	TSVMod Method	MSALL	MSALL
11	TSVMod RMSD	12.889	16.145
12	TSVMod NO35	0.096	0.031

Table :1 (C) Homology modelling of PRSV-P & PRSV-W CP using Swiss Model

Sr.No	Global Quality estimate	PRSV-P	PRSV-W
1	QMEAN	-3.55	-3.35
2	CB	-2.11	-1.86
3	All Atom	-2.80	-2.82
4	Solvation	-1.26	-1.15
5	Torsion	-2.73	-2.59
6	Template	5odv.1.A	5odv.1.A
7	Sequence identity	62.61%	57.66%

Table :2 Ab initio modelling of PLCV using I-TASSER

PLCV	Software	Overall quality	Ramachandran score	Percent Residue in most favoured region	Percent Phi/Psi pairs in disallowed region
	I-TASSER	1.5	6.69	52%	15.56

Table: 3 Physicochemical characterization of coat-protein by Expasy's Protparam tool

Virus Name	Molecular weight (Da)	pI	-R	+R	EC	II	AI	GRAVY
PLCV	30134.77	10.0	21	45	44600	46.98	67.30	-0.590
PapMV	23045.14	7.73	15	16	17085	48.81	74.09	-0.124
PRSV-P	33128.13	7.76	43	44	29910	29.00	61.42	-0.933
PRSV-W	33047.13	7.77	44	45	35410	29.97	61.85	-0.943

Table 4. Physicochemical characterization of coat-protein by SOPMA

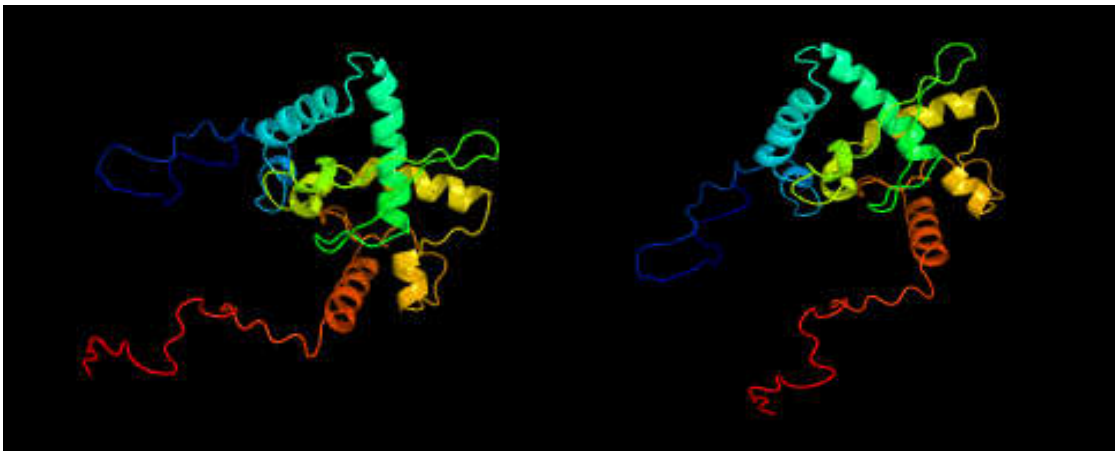
Parameters	PLCV	PapMV	PRSV-P	PRSV-W
Alpha helix	20.08%	42.79%	40.83%	43.55%
3_{10} helix	0.00%	0.00%	0.00%	0.00%
Pi helix	0.00%	0.00%	0.00%	0.00%
Beta bridge	0.00%	0.00%	0.00%	0.00%
Extended strand	25.10%	8.37%	11.07%	11.50%
Beta turn	5.41%	4.65%	4.50%	3.48%
Bend region	0.00%	0.00%	0.00%	0.00%
Random coil	49.42%	44.19%	43.60%	41.46%
Ambiguous states	0.00%	0.00%	0.00%	0.00%
Other states	0.00%	0.00%	0.00%	0.00%

Table 5 : PRSV-P & PRSV-W Homology model validation by PROSESS tool

	Software	Overall quality	Ramachandran score	Percent Residue in most favoured region	Percent Phi/Psi pairs in disallowed region
PRSV-P	Phyre2	3.5	0.52	79%	0.0
	ModWeb	4.5	0.64	91%	0.0
	Swiss model	4.5	0.37	84%	0.0
PRSV-W	Phyre2	3.5	0.54	80%	0.0
	ModWeb	4.5	0.48	88%	0.73
	Swiss model	4.5	0.35	85%	0.0

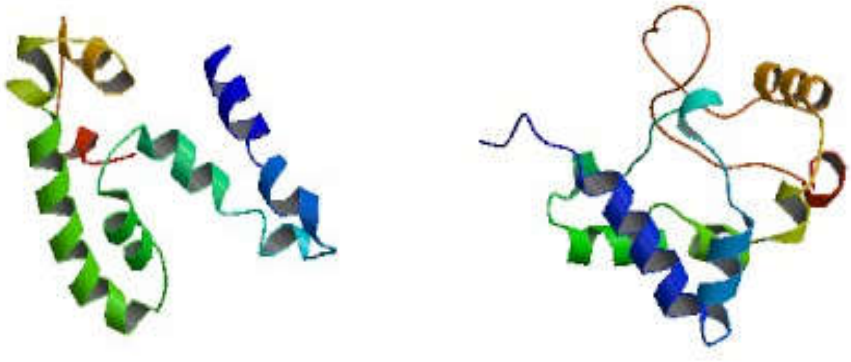
Table :6 PLCV Ab initio model validation by PROSESS tool

	Software	Overall quality	Ramachandran score	Percent Residue in most favoured region	Percent Phi/Psi pairs in disallowed region
PLCV	I-TASSER	1.5	6.69	52%	15.56



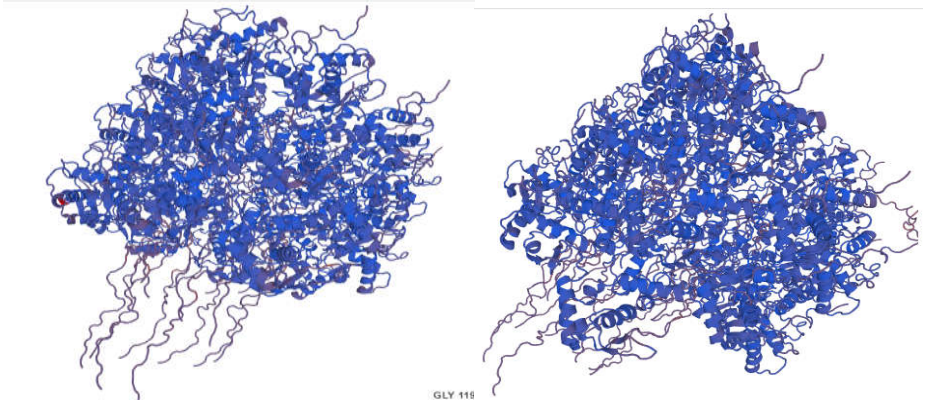
C5odvB_ template of PRSV-P C5odvB_ template of PRSV-W

Fig.: 2 (A) Homology modelling of PRSV-P & PRSV-W CP using PHYRE2



template of PRSV-P template of PRSV-W

Fig :2 (B) Homology modelling of PRSV-P & PRSV-W CP using ModWeb



template of PRSV-P template of PRSV-W

Fig :2 (C) Homology modelling of PRSV-P & PRSV-W CP using Swiss Model

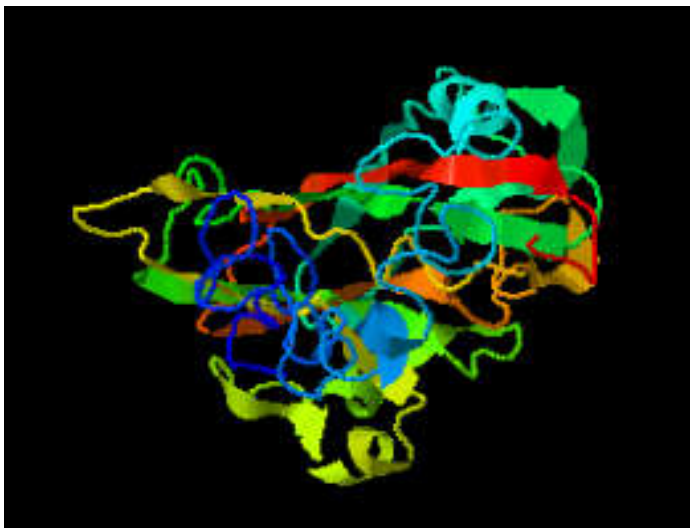


Fig.: 3 Ab initio modelling of PLCV using I-TASSER

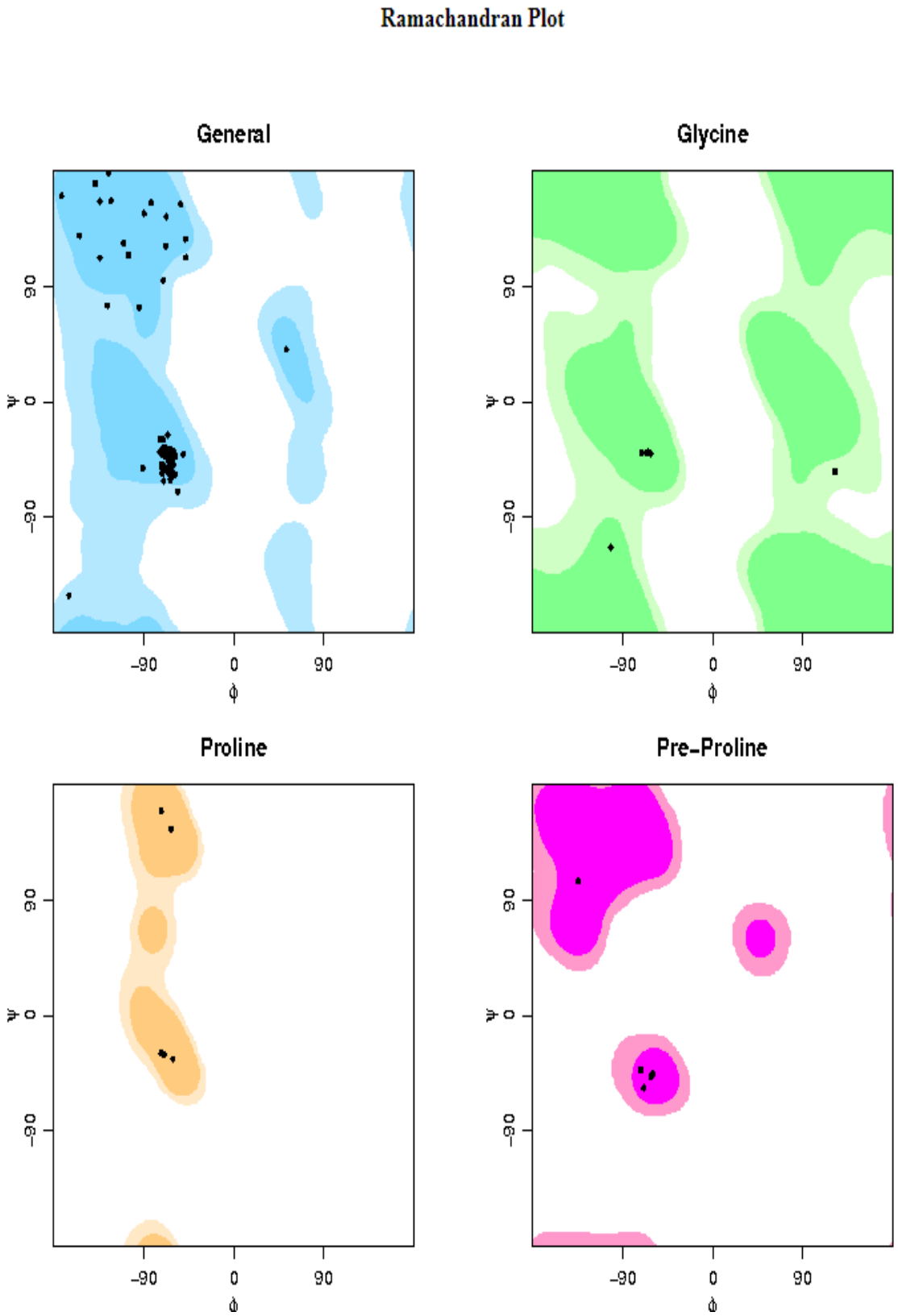


Fig: 4 (A) Ramachandran Plot of PRSV-P built model by Modweb tool

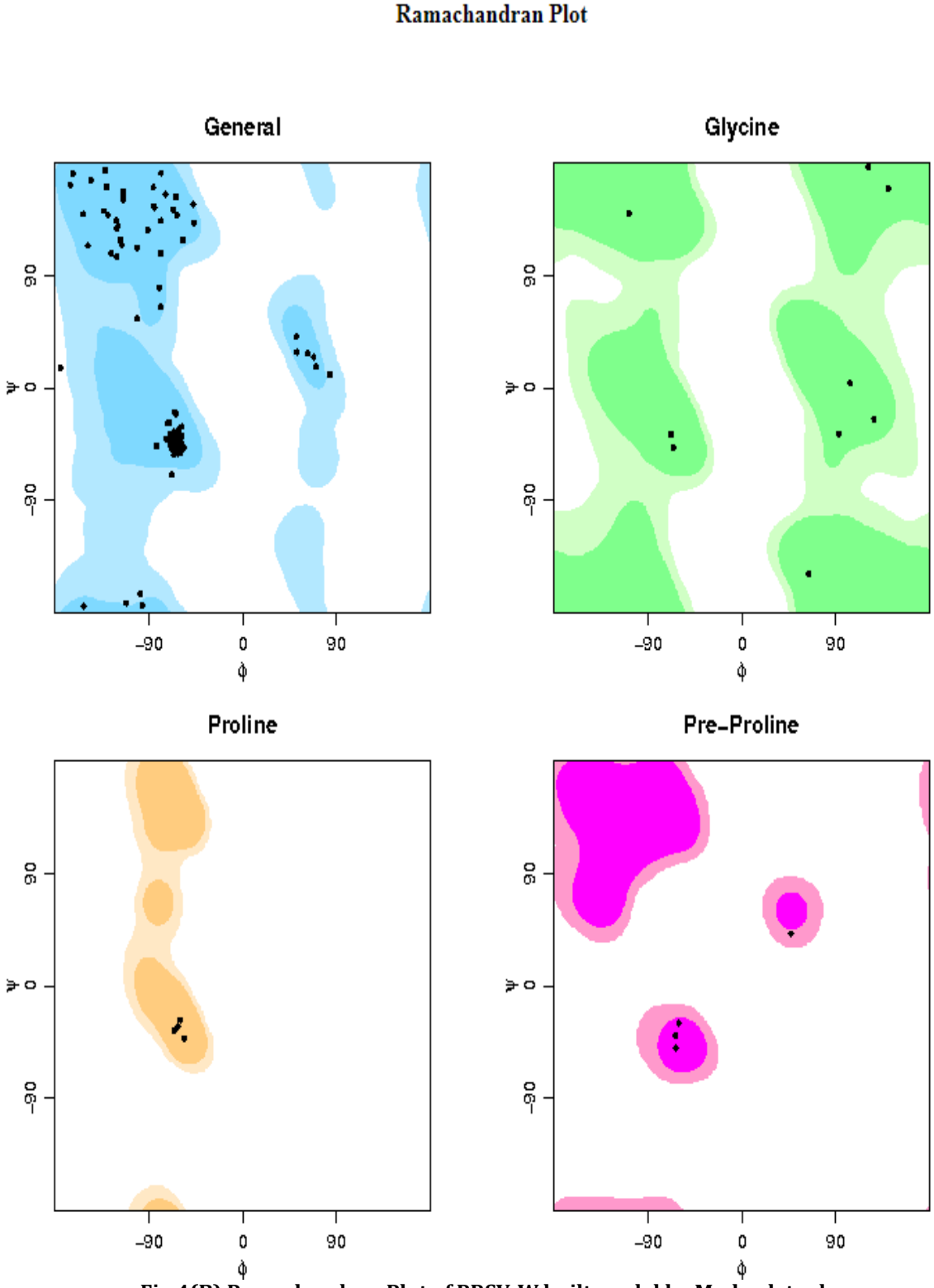


Fig:4(B) Ramachandran Plot of PRSV-W built model by Modweb tool

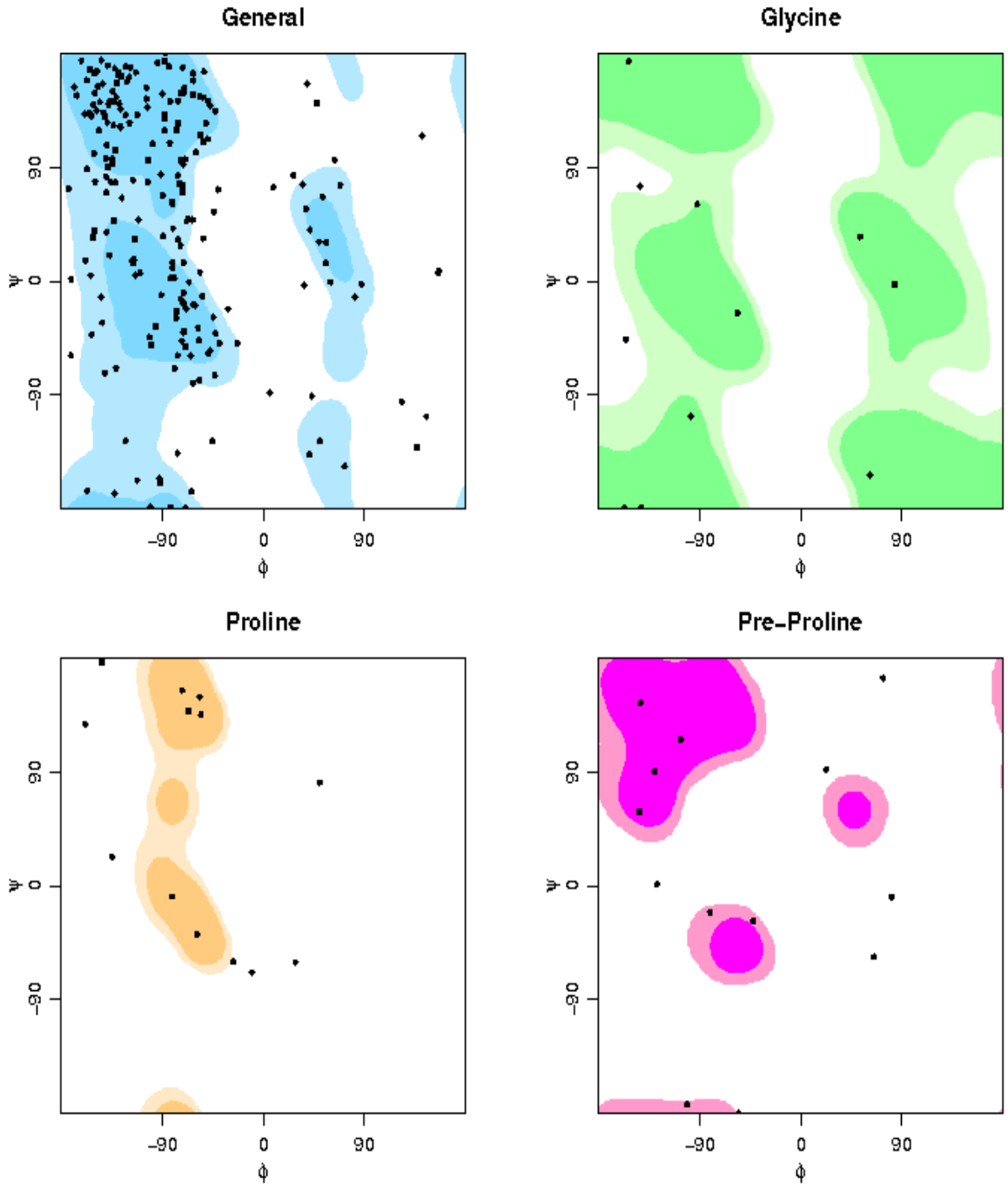


Fig: 5 Ramachandran plot of PLCV built model by I-TASSER tool

CONCLUSION

Structural bioinformatics is concerned with computational approaches to predict and analyze the structure and function of protein.In this study, CP of PRSV-P, PRSV-Wand PLCV were selected and PapMV taken as reference. Comparative homology modelling & Ab initio were performed by different servers, viz. ModWeb, Phyre2,Swiss Model & I-TASSER with the observation that the model generated by modweb for both PRSV-P & PRSV-W has most accurate, of high quality and acceptable highest percent residue in most favoured region 91% and 81% respectively. While PLCV, over all quality was 1.5 in I-TASSER with 52% percent residue in most favoured region. The values of GRAVY were -0.590, -0.124, -0.933 & -0.943

PLCV, PapMV, PRSV-P & PRSV-W respectively indicating the possibility of better interaction with water. The models were validated using protein structure checking tools PROCESS. These structures will provide a good foundation for functional analysis of experimentally derived crystal structures.The study may help in understanding the protein Structure, function, number and types of epitopes, immunogenic portions, and suitability for antibody production, taxonomic studies, evaluation studies and virus diagnostics.

REFERENCES

1. Chalak, S.U., Hasbnis, S.N. and Supe, V.S. (2017) Papaya ring spot disease management: Rev. J. Pharmacognosy and Phytochemistry ; 6(5): 1911-1914.
2. Lokhande, N.M., Moghe, P.G., Matte, A.D. and Hiware, B.J. (1992). Occurrence of *Papaya ringspot virus* (PRSV) in Vidarbha regions of Maharashtra. J. Soils and Crops 2:36-39.
3. Gonsalves, D. (1998) Control of Papaya Ringspot Virus in Papaya: A Case Study , Annu. Rev. Phytopathol. 36:415–37
4. Yeh, S.D. and Gonsalves, D. (1984) Evaluation of induced mutants of papaya ringspot virus for control by cross protection, Phytopathol, 74:1086-1091.
5. Chavan, V.M., Tomar, S.P.S. and Dhale, M.G. (2010) Management of Papaya Ring Spot Virus (PRSV-P) of Papaya under Pune Conditions, Proc. IInd IS on Papaya Acta Hort. 851,pp447-451.
6. Yang, S., Wang, T, Bohon, J., Gagné, M-È, Bolduc, M. , Leclerc, D. and Li, H. (2012) Crystal structure of the coat protein of the flexible filamentous papaya mosaic virus, Mol Biol. 14; 422(2): 263–273
7. Yadav, N. K., Sarika-Iquebal, M.A. and Akram, M (2011) In-silico analysis and homology modelling of coat-protein of Mungbean Yellow Mosaic India Virus J. Food Legumes, 24(2): 138-141.
8. Prajapati, C. and Bhagat, C (2012) In-Silicoanalysis and homology modeling of target Proteins For *Clostridium botulinum*. IJPSR, 3(7) 2050-2056.
9. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005). Protein Identification and Analysis Tools onExPASy Server. In: John M. Walker (Ed), The Proteomics Protocols Handbook, Humana Press. Pp 571-607.
10. Sahay, A. and Shakya, M (2010) In silico Analysis and Homology Modelling of Antioxidant Proteins of Spinach, J Proteomics Bioinformatics, 3(5):148-154
11. Morris, A. L., MacArthur, M.W., Hutchinson, E.G. and Thornton , J.M.(1992) Stereochemical quality of protein structure coordinates, Proteins, 12 :345-364
12. Gava, D., Serrão, V.H.B. , Fernandes, L,T, Cantão, M. E., ReisCiacci-Zanella, J., Morés, N. and Schaefer,R.(2018) Structure analysis of capsid protein of Porcine circovirus type 2 from pigs with systemic disease, Brazilian J. Microbiol, 49(2):351-357

CITATION OF THIS ARTICLE

Jay Patel and Rishee K. Kalaria. *In-silico* Analysis & Homology modelling of Papaya Ring Spot Virus & Papaya Leaf Curl virus coat protein. Bull. Env. Pharmacol. Life Sci., Vol 7 [8] July 2018 : 26-35.