www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; 11(4): 516-521

www.thepharmajournal.com
Received: 02-01-2022
Accepted: 09-02-2022

R Anantapur
Ph.D., Scholar, Department of Plant Biotechnology, UAS, GKVK, Bengaluru, Karnataka, India

R Mohan
Ph.D., Scholar, Department of Plant Biotechnology, UAS, GKVK, Bengaluru, Karnataka, India

A Peter
Professor, Department of Plant Biotechnology, UAS, GKVK, Bengaluru, Karnataka, India

SG Bhat
Research Scholar, College of Forestry, Sirsi, UAS, Dharwad, Karnataka, India

Kiran NR
Research Scholar, CMAP, Bengaluru, Karnataka, India

Corresponding Author:
R Anantapur
Ph.D., Scholar, Department of Plant Biotechnology, UAS, GKVK, Bengaluru, Karnataka, India

The 3D structure prediction, quality assessment and model validation of Replicase (Rep) protein of mungbean yellow mosaic virus (MYMV)

R Anantapur, R Mohan, A Peter, SG Bhat and Kiran NR

Abstract
Mungbean yellow mosaic virus (MYMV) is a devastating virus in green gram cultivation. The Replicase initiation (Rep) / AC1 protein is encoded by *Rep* gene from DNA A of bipartite viral genome of MYMV. The Rep protein is important among all as it is involved in replication and multiplication of virus in host system. The 3D structure prediction, quality assessment and model validation were done by different online molecular web tools. The 3D structure was developed *Ab-initio*, simultaneously refinement of the model was done, followed by assessment of the quality of the Rep protein by studying the stereo chemical properties. Physiochemical properties and amino acid content were also studied.

Keywords: Replicase, replication, assessment, validation, refinement

1. Introduction
Mung bean is important short duration rainfed legume crop of Indian. It is mostly cultivated throughout Indian sub-continent and also parts of South east and East Asia regions. The crop productivity of mungbean is affected by several challenges in the form of abiotic and biotic stresses. The impact of biotic stresses like various pest and diseases is enormous and among which the viral diseases like mungbean yellow mosaic virus (MYMV) cause devastation to all the mungbean growing areas (Borah and Dasgupta, 2012) [2]. MYMV is single stranded bipartite DNA virus belonging to the genus *Geminivirus*, family: *Begomoviridae* and transmitted in a circulative persistent manner by insect vector, whitefly (*Bemisia tabaci*), (Ilyas *et al*., 2009) [11]. The control of the disease can be done by management of virulent insect vectors, which is an indirect and less promising. Mungbean virus has bipartite genome consisting of components, DNA A and DNA B (Qazi *et al*., 2007) [18] (Fig.1a). Rep or AC1 protein is a part of DNA A which is of 40kD and codes for replication initiation. Geminiviruses encode only a few proteins for their replication and recruit most of their replication machinery from their plant hosts (Hanley *et al*., 1999) [9]. They replicate by rolling circle mode of replication (RCR) and the viral Rep protein initiates RCR by the site-specific nicking at a conserved nonamer (TAATATTAC) sequence (Raghavan *et al*.,2004) [19]. Rep is a multifunctional protein having site-specific nicking and ligation, site-specific DNA-binding and ATPase activities (Elmer *et al*., 1988, Pant *et al*., 2001, Raghavan *et al*., 2004) [7, 17, 19]. It regulates its own expression at transcriptional level and is also known to induce host replication machinery, presumably to enable the virus to replicate in differentiated cells (Eglekrout *et al.*, 2002) [6]. The N-terminal part of Rep largely contains DNA-binding, nicking-ligation and oligomerization domains, while the C-terminal half contains ATP binding and ATPase activity domains (Choudhury *et al.*, 2006) [3]. (Fig. 1b) Rep binds to specific sequences (repeats and introns) present in the catalytic region and hydrolyses the phosphodiester bond between the seventh and eighth residues of the invariant nonamer 5' TAATATT↓AC 3' (Laufs *et al.*, 1995) [16]. Rep remains bound covalently to the 5' phosphate end, and the 3' hydroxyl end thus generated becomes available for rolling-circle replication. After a full cycle of replication, the new origin sequence is generated, which is again hydrolysed by Rep. Subsequently Rep ligates the nascent 3' end of DNA with the previously generated 5' end. In this way, a unit genome length, circular, single-stranded DNA molecule, the mature viral genome, is processed. Once the infected whitefly feeds on the healthy leaf sap, the virions are transmitted to the phloem-associated cells and the virus initiates infection on the healthy plants. After its entry, virion releases ssDNA and becomes dsDNA by using hosts DNA polymerase (Saunders *et al*., 1991) [21].

Later, dsDNA is transcribed to produce Replication associated protein (Rep) by using host RNA polymerase II which is necessary for viral replication. This protein initiates viral replication, which occurs by a combination of rolling-circle replication (Saunders *et al*., 1991; Stenger *et al*., 1991) [21, 24] from double stranded (ds) intermediate and recombination-dependent replication (Jeske *et al*., 2001) [12]. During evolutionary process these diversity patterns of viral genes are presumably being driven by introductions, isolation, geographical distribution and disease management strategies (Qazi *et al*., 2007) [18]. In the different geographic regions, a significant difference was observed at the nucleotide level, but the amino acid sequences were quite conserved. It is thought that during distinct evolutionary pathways in different geographic origin, the difference in nucleotide similarities could have occurred due to wobble bases of codons and thus amino acid residues were conserved suggesting a common selection pressure to keep the same protein confirmation (Wang *et al*., 1994) [25]. Based on the importance of the Rep protein in viral replication and its conserved common amino acid residue sequence in MYMV diversified viral genome, the Rep protein was selected for Ab-initio development of its 3D protein model by available sequences. This model will help in assessment and molecular docking studies of the protein, in turn leading to detect the suitable binding antiviral compounds within the host or natural available compounds or synthetic products before an on filed studies.

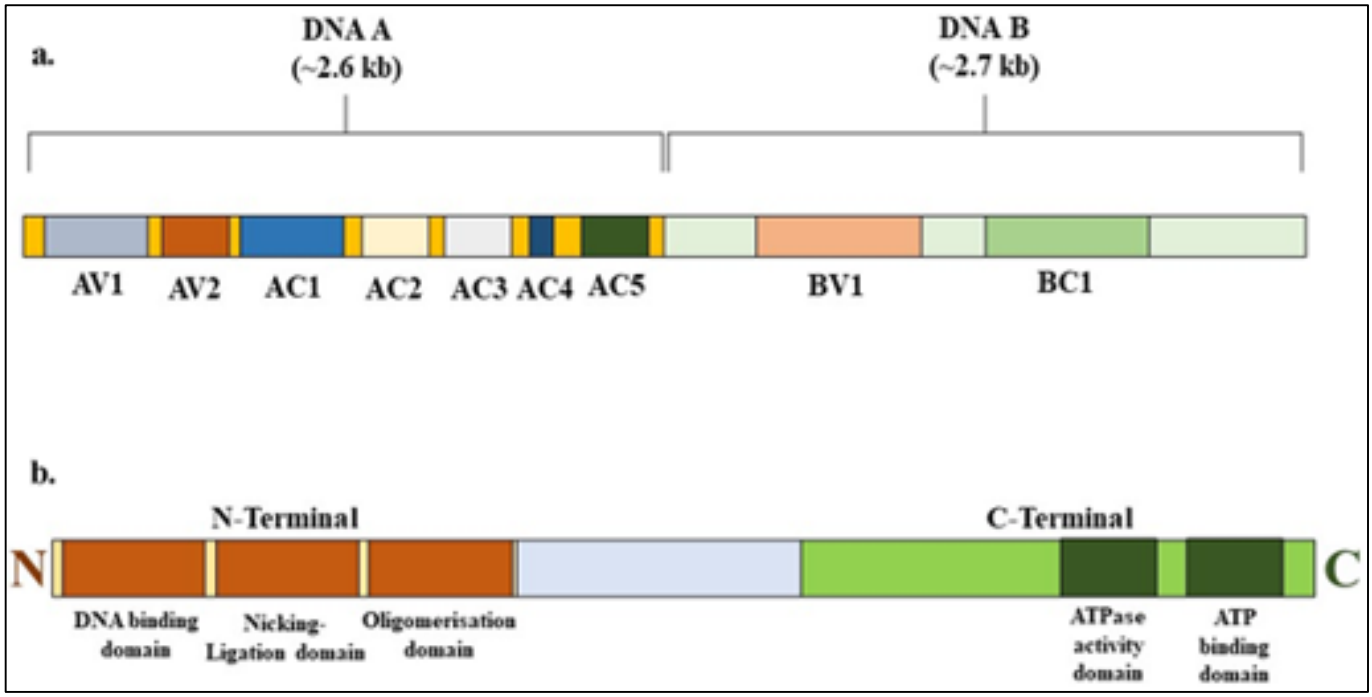

Fig 1: a. Diagrammatic representation of bipartite genome of MYMV, b. Diagrammatic representation of the components of Rep protein

2. Material and Methods

2.1. Web tools

The ab-initio modeling of the Rep protein requires many of the online based molecular web tools, which can be used to design, asses and develop the final model of Rep protein, the tools used are as follows-

Sl. No.	Web-tool	Purpose
1.	NCBI database (https://www.ncbi.nlm.nih.gov/)	Retrieval of sequences
2.	ExPASy Protparam server (https://web.expasy.org/protparam/)	Physiochemical characteristics of protein sequence
3.	I-TASSER, version 5.1 (https://zhanggroup.org/I-TASSER/)	*Ab initio* prediction of 3D structure of protein
4.	GalaxyRefine 2 webserver (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE)	Refinement of 3D structure
5.	SAVES 6.0 (https://saves.mbi.ucla.edu/)	Validation of 3D structure

2.2. 3D structure prediction and development

Iterative Threading Assembly Refinement (I-TASSER, version 5.1) was used to predict the 3D structure of the Rep protein. I-TASSER is an automated server and a hierarchical protein structure modeling approach based on the statistical significance of the profile-profile alignment (PPA) threading alignments and the structure convergence of the Monte Carlo simulations (Zhang and Skolnick, 2004; Wu and Zhang, 2007) [27, 26]. In brief, the output file comprises a predicted confidence score (C-score) for each of the five models in the range of -2 to +5, which predicts the reliability and the quality of the model, with the RMSD and TM score for the first model (Wu *et al*., 2007; Zhang, 2008) [26, 28]. The models with the lowest confidence score (C-score) were subsequently selected for the structural refinement using an *Ab initio* loop and terminus-based GalaxyRefine 2 webserver (Ko *et al*., 2012) [14].

1.3. Quality assessment and model validation

The quality of the final model was assessed by the PROCHECK program of Structural Analysis and Verification (SAVES 6.0) tool to determine the stereo chemical quality of the model by generating a Ramachandran plot (Laskowski *et al*., 1993; Sivaramakrishnan *et al*., 2012; Sidhu *et al*., 2020) [15, 23, 22].

3. Results and Discussion

3.1. 3D Structure prediction of MYMV Rep protein

The 3D structure was determined by the I-TASSER server. It generates structural confirmations, then uses the SPICKER clustering program to cluster all structures based on pairwise structural similarity (Zhang, 2008) [28]. The model with the lowest C-score value relative to the native was selected as the best model (Fig. 2) The C-score, TM score and RMSD value of MYMV Rep protein were -1.12, 0.57±0.14 and 9.1±4.6 Å respectively. The top predicted model was submitted to the Galaxy Refine 2 server for refinement. It rebuilds the side-chain and performs side-chain repacking and structure relaxation by molecular dynamic simulation (Ko *et al*., 2012) [14].

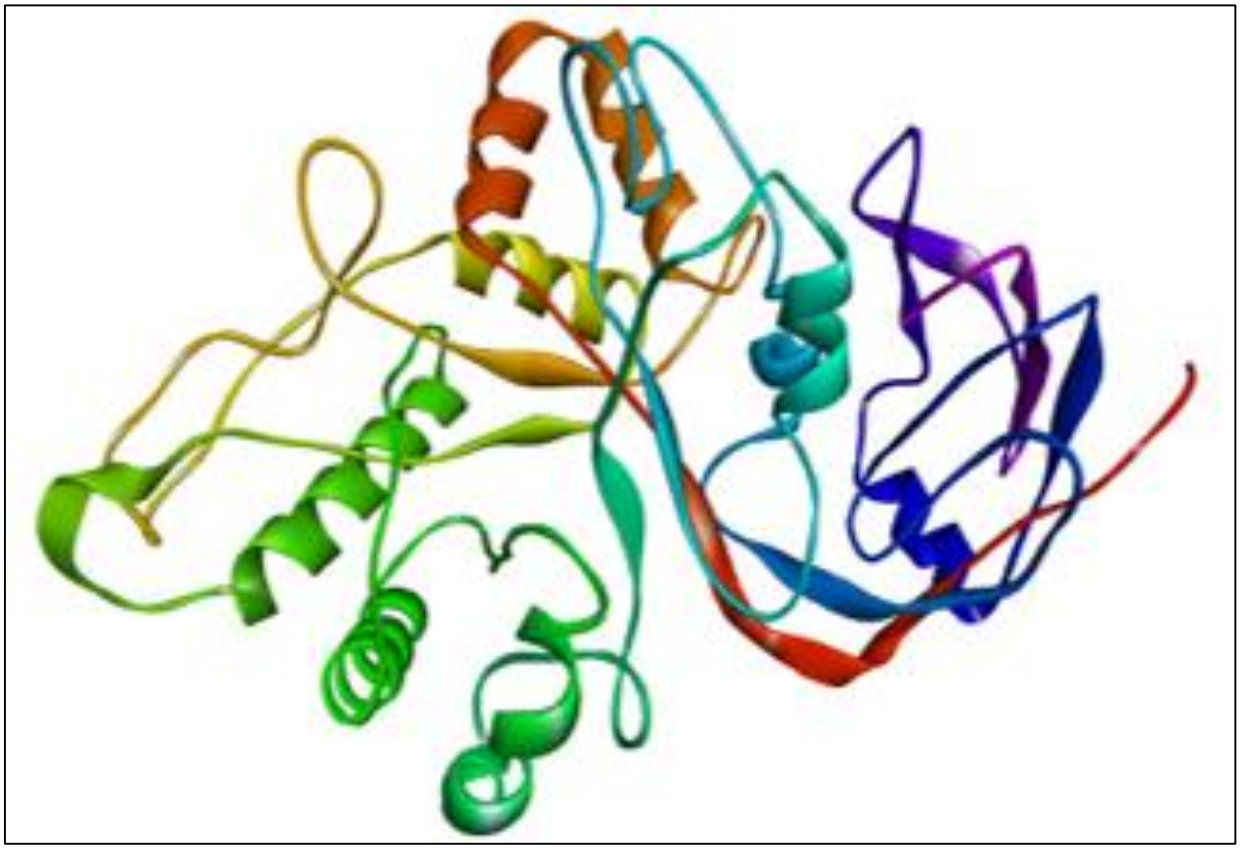

Fig 2: Three-dimensional cartoon model representing MYMV Rep protein (PyMol)

3.2. Quality assessment and model validation of MYMV Rep protein

3.2.1. Stereo chemical properties

To validate the stereo chemical quality of the 3D models in each step of refinement, they were uploaded to SAVES online tool (Daróczi, 2013) [4]. to obtain the Ramachandran plot in PROCHECK. The models with good quality were visualized in Biovia Discovery Studio 2020 (BIOVIA, 2020) [1] and PyMOL (DeLano, 2009) [5]. The Ramachandran plot provides information regarding the torsional angles - phi (φ) and psi (ψ) - of the residues (amino acids) contained in a protein. The Ramachandran plot revealed that, most of the amino acid residues of MYMV Rep protein was in the favored and allowed regions (97.1 per cent) with 2.8 per cent residues in the disallowed region respectively indicating a good quality of the refined models (Fig. 3). Moreover, the protein had the majority of the amino acid residues favoring β sheets and right-hand α helix.

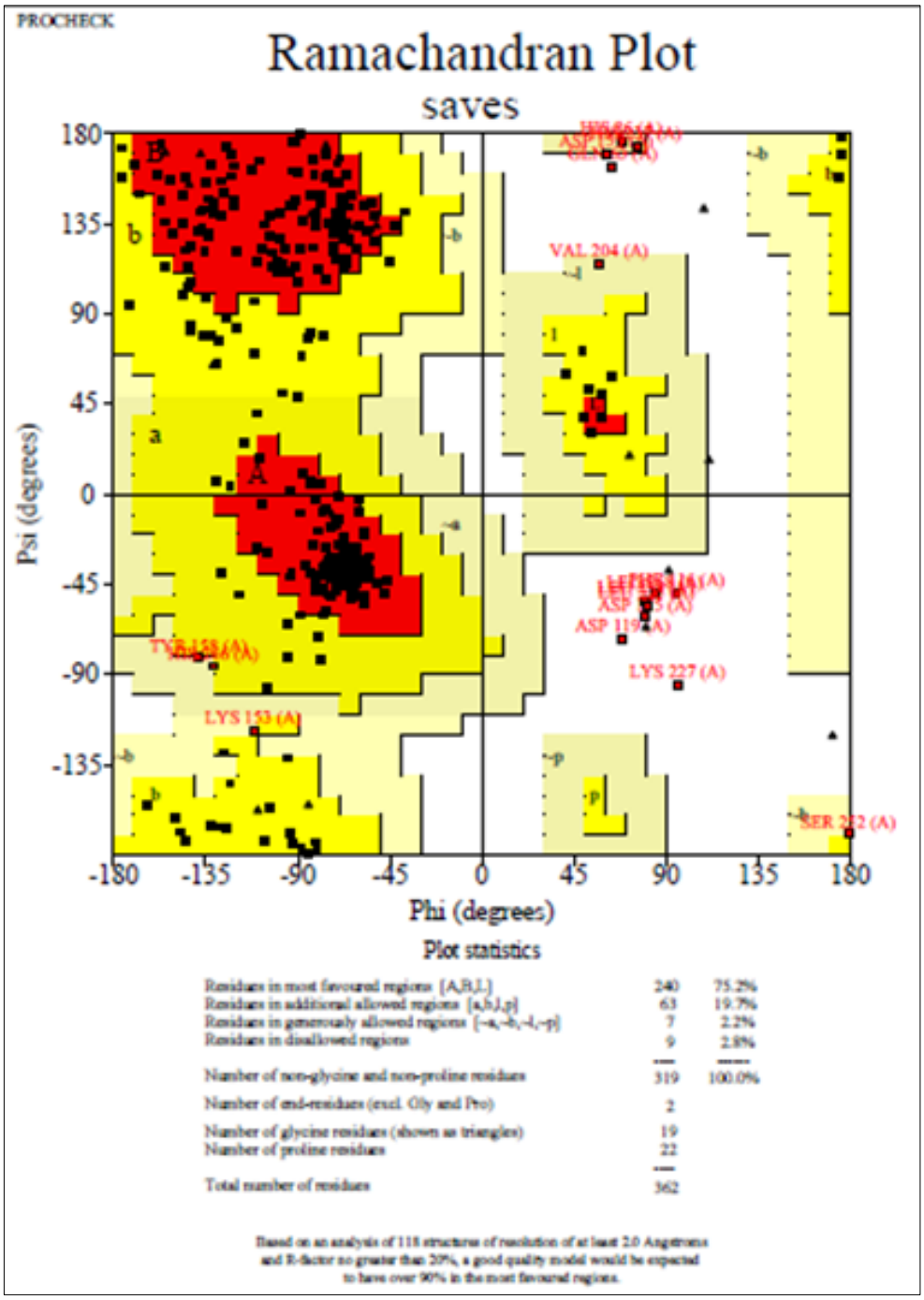

Fig 3: Ramachandran plot validation for MYMV Rep protein generated by PROCHECK program of SAVES 6.0 tool

3.2.2. Physicochemical properties: Physicochemical properties are considered to be essential factors for predicting the function and structure of protein sequences. The primary sequence analysis computed by Protparam and the estimated parameters for MYMV Rep protein is summarized in Table 1. The Mw of MYMV Rep protein is 40.83 kDa and theoretical pI was 7.18, the isoelectric point (pI) is the pH at which the protein carries zero electric charge, the pI reveals that the protein carries no electric charge at 10.06 pH and this value plays a significant role in protein purification as the solubility is minimal, protein is stable and compact (Sahay *et al*., 2010) [20]. The amino acid composition of MYMV Rep protein computed by Protparam is given in Fig. 4. The extinction coefficient value at 280 nm was 53205 M^{-1} cm^{-1}. The extinction coefficient helps in the quantitative study of protein-protein and protein-ligand interactions (Jethra *et al.*, 2012) [13]. The analysis showed that the insatiability index (II) was 40.25 and it is predicted that if II is more than 40, the protein is probably unstable; any value less than 40 indicates protein stability. The aliphatic index (AI) was 74.14, higher value of AI plays an important role in thermal stability of protein (Ikai, 1980) [10]. The MYMV Rep protein had negative Grand average of hydropathicity (GRAVY) value showing the hydrophilicity due to its acidic nature (Gasteiger *et al.*, 2005) [8] and it calculates the hydropathic value of amino acid residues. The atomic composition of Rep protein is, Carbon: 1821, Hydrogen: 2799, Nitrogen: 505, Oxygen: 547 and Sulfur: 10. Total number of atoms are 5682 and estimated half-life as the N-terminal of the sequence considered is M (Met) estimated half-life is 30 hours (mammalian reticulocytes, *in vitro*), >20 hours (yeast, *in vivo*) and >10 hours (*Escherichia coli*, *in vivo*).

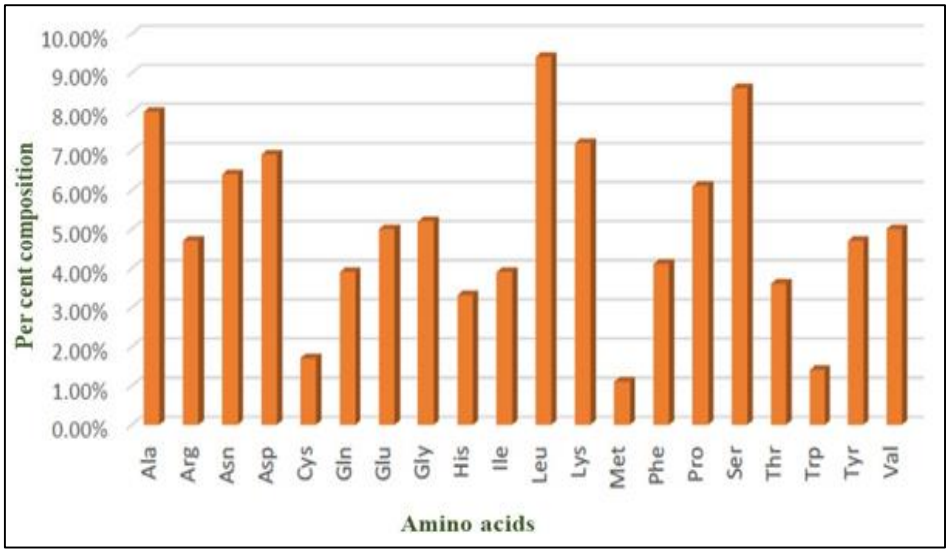

Fig 4: Amino acid composition of MYMV Rep protein as computed by Protparam

Table 1: Physio-chemical parameters of the MYMV Rep protein computed by Protparam

Sl. No.	Computed parameters	MYMV Rep protein
1	Number of amino acids	362
2	Molecular weight (kilo Daltons)	40083
3	Theoretical pI	7.18
4	Total number of negatively charged residues (Asp + Glu)	43
5	Total number of positively charged residues (Arg + Lys)	43
6	The instability index	40.25
7	Aliphatic index	74.14
8	Grand average of hydropathicity (GRAVY)	-0.594

4. Conclusion

The Rep protein model prediction, assessment and validation provide the 3D protein structure which can be utilized for the binding sites prediction which help in molecular docking. The stereochemical properties and physiochemical properties in datil studied will describe the amino acid composition of the protein, molecular weight, isoelectric point, instability index, aliphatic index, GRAVY score, half-life in detail, these details help in the research aspects of protein-ligand or protein-protein interactions of the MYMV Rep protein during virus-host interactions. The study will support the *in-silico* analysis of the antiviral compounds of MYMV and related virus.

5. Acknowledgement

The authors are thankful to the Department of Plant Biotechnology, UAS, Bangalore, for providing all the facilities to conduct the research work and also acknowledge Karnataka Science and Technology Promoting Society (KSTePS), Govt. of Karnataka for providing DST-PhD fellowship.

6. References

1. Biovia, Biovia Workbook, Release BIOVIA Pipeline Pilot. 2017.
2. Borah BK, Dasgupta I. Begomovirus research in India: A critical appraisal and the way ahead. J. Biosci. 2012;37(4):791-806.
3. Choudhury NR, Malik PS, Singh DK, Islam MN, Kaliappan K, Mukherjee SK. The oligomeric Rep protein of Mungbean yellow mosaic India virus (MYMIV) is a likely replicative helicase. Nuc. Acids Res. 2006;34(21):6362-6377.
4. Daróczi G. Saves: Fast load variables. R package version 0.5, 2013. http://cran.r-project.org/package=saves.
5. Delano. The PyMOL Molecular Graphics System. 2009. https://pymol.org/2/.
6. Egelkrout EM, Mariconti L, Settlage SB, Cella R, Robertson D, Hanley-Bowdoin L. Two E2F elements regulate the proliferating cell nuclear antigen promoter differently during leaf development. Plant Cell. 2002;(14):3225-3236.
7. Elmer JS, Brand L, Sunter G, Gardiner WE, Bisaro DM, Rogers SG. Genetic analysis of the tomato golden mosaic virus. II. The product of the AL1 coding sequence is required for replication. Nucleic Acids Res. 1988;(16):7043-7060.
8. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook. 2005, 571-607pp.
9. Hanley-Bowdoin L, Settlage SB, Orozco BM, Nagar S, Robertson D. Geminiviruses: models for plant DNA replication, transcription, and cell cycle regulation. Critical Rev. in Plant Sci. 1999;18(1):71-106.
10. Ikai A. Thermo-stability and aliphatic index of globular proteins. J. Biochem. 88(6):1895-1898.
11. Ilyas M, Qazi J, Mansoor S, Briddon, RW. Molecular characterization and infectivity of a "Legumovirus" (genus Begomovirus: family Geminiviridae) infecting the leguminous weed *Rhynchosia minima* in Pakistan. Virus Res. 1980;145:279-284.
12. Jeske H, Lutgemeier M, Preiss W. DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. EMBOJ. 2001;(20):6158-6167.

13. Jethra G, Mishra AK, Pandey PS, Chandrasekharan H. Structure and function prediction of unknown wheat protein using LOMETS and I-TASSER. Indian J Agric. Sci. 2012;82(10):867.
14. Ko J, Park H, Heo L, Seok C. Galaxy WEB server for protein structure prediction and refinement. Nucleic Acids Res. 2012;40(1):294-297.
15. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J Appl. Crystallogr. 1993;26(2):283-291.
16. Laufs J, Schumacher S, Geisler N, Jupin I, Gronenborn B. Identification of the nicking tyrosine of geminivirus Rep protein. FEBS Letters. 1995;377:258-262.
17. Pant V, Gupta D, Roy Choudhury N, Malathi VG, Varma A, Mukherjee SK. Molecular characterization of the Rep protein of the blackgram isolate of Indian mungbean yellow mosaic virus. J Gen. Virol. 2001;(82):2559-2567.
18. Qazi J, Ilyas M, Mansoor S, Briddon RW. Legume yellow mosaic viruses: genetically isolated begomoviruses. Mol. Plant Pathol. 2007;8(4):343-348.
19. Raghavan V, Malik PS, Choudhury NR, Mukherjee SK. The DNA-A component of a plant geminivirus (Indian mung bean yellow mosaic virus) replicates in budding cells. J. Virol. 2004;(78):2405-2413.
20. Sahay A, Shakya M. In silico analysis and homology modelling of antioxidant proteins of spinach. J Proteomics Bioinform.,2010;3:148-154.
21. Saunders K, Lucy A, Stanley J. DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. Nuc. Acids Res. 1991;19:2325-2330.
22. Sidhu KS, Bhangu SK, Pathak RK, Yadav IS, Chhuneja P. Identification of natural lead compounds for leaf rust of Wheat: A molecular docking and simulation study. J Prot. Proteomics. 2020;11(4):283–295.
23. Sivaramakrishnan V, Thiyagarajan C, Kalaivanan S, Selvakumar R, Anusuyadevi M, Jayachandran KS. Homology modeling, molecular docking and electrostatic potential analysis of MurF ligase from *Klebsiella pneumonia*. Bioinformation. 2012;8(10):466–473.
24. Stenger DC, Revington GN, Stevenson MC, Bisaro DM. Replication release of geminivirus genomes from tandemly repeated copies: Evidence for rolling-circle replication of a plant viral DNA. Proc. Nat. Acad. Sci. 1991;88:8029-8033.
25. Wang X, Eggenberger AL, Nutter JRFW, Hill JH. Pathogen-derived transgenic resistance to soybean mosaic virus in soybean. Mol. Breed. 1994;8:119-127.
26. Wu ST, Zhang Y. Lomets: A local meta-threading-server for protein structure prediction. Nucleic Acids Res. 2007;10:3375-3382.
27. Zhang Y, Skolnick J. Automated structure prediction of weakly homologous proteins on a genomic scale. PNAS. 2004;101(20):7594-7599.
28. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008;9(1):1-8.