In silico: Coat Protein of PepYLCV-APWS

T Runifah¹, B Nova², D H Tjong¹,³, and J Jamsari¹,⁴

¹Department of Biotechnology, Postgraduate Program, Andalas University, Padang, West Sumatera, 25163 Indonesia
²Agricultural Sciences Department, Doctoral Program, Andalas University, Padang, West Sumatera, 25163 Indonesia
³Department of Biology, Mathematics and Life Sciences Faculty, Andalas University, Padang, West Sumatera, 25163 Indonesia
⁴Department of Agrotechnology, Agriculture Faculty, Andalas University, Padang, West Sumatera, 25163 Indonesia

Email: jamsari@agr.unand.ac.id

Abstract. Pepper Yellow Leaf Curl Diseases [PepYLCD] one of the most damaging diseases in pepper and chili caused by a geminivirus. The coat protein of geminivirus plays an important role in encapsidation of the virus genome and also as a transmission mediator. This research aimed to predict the 3D structure and binding site of coat protein pepper leaf curl virus isolate Alahan Panjang, West Sumatera [PepYLCV-APWS]. The prediction was using swiss-model and chimera software for protein modeling. Furthermore, this study will be helpful to understand the similarity and the molecular-interaction to develop robust resistant strategies against geminiviruses in the nearby future.

Keywords: binding site, coat protein, geminivirus, in silico, PepYLCV

1. Introduction
Pepper yellow leaf curl virus [PepYLCV] is one of the most damaging geminiviruses in pepper and chili crop cultivation. They are transmitted by the whitefly, Bemisia tabacii, when the insects suck the sap from plant leaves. The viral genome released from the coat protein [capsid] enters the cytoplasm and continues into the nucleus when the virus attaches to the cell wall [1], [2]. In the nucleus, viral DNA is converted into double-stranded DNA from single-stranded DNA by the host plant DNA polymerase. Then dsDNA is transcribed to synthesize Rep for viral DNA replication. Through the Rolling-circle Replication [RCR] or Recombination Dependent Replication mechanism, this replication will produce ssDNA replication which will be converted back into dsDNA for multiplication and virulence protein synthesis in cells, and ssDNA encapsulated (fusion of virions) by coat protein [CP] into new virions [3], [4], [5], [6].

New virion particles will be transmitted to other cells or into the phloem via the plasmodesmata for systemic dissemination with the help of Movement Protein [MP] and Nuclear Shuttle Protein [NSP] in the capsid. Systemic spread/dissemination of the virus is the spread of viral infection to host cells from distal to the site of inoculation to all cells of other organs [4], [6], [7]. CP, which functions as a wrapper for the viral genome [capsid] and as a vector transmission mediator or to determine the specificity of insect vectors. Besides, CP also functions as an NSP in monopartite viruses [1], [2], [8]. In this study, we characterized the PepYLCV coat protein from Alahan Panjang West Sumatera...
through a computational approach to see the functional structure and protein folding. These results are expected to help understand the virus and thus provide a way to control the virus.

2. Materials and Method

2.1. DNA and Protein Sequence
The PepYLCV-APWS AV1 DNA sequence was obtained from Plant Virus Genome Collection, Laboratory of Biotechnology, Agriculture Faculty, Andalas University. The DNA sequence was translated into a protein sequence using the BioEdit 7.2.5 software [9].

2.2. In Silico Protein Characterization and Homology Protein Fold Modelling
The protein sequence was checked for the presence of a conserved domain using NCBI-CDD web tools [10], [11]. The 3D structure of Coat Protein PepYLCV-APWS was predicted using Swiss-Model web software [12]. Quality assessment of the predicted structure was performed with ProFunc to generate The Ramachandran plot [13].

2.3. Domain and Binding Site Prediction
CASTp binding site was used to predict voids and pockets in the predicted structure and determine possible and potential binding sites [14].

2.4. In Silico Protein Visualization
USCF Chimera 1.15 was performed to visualize protein modeling and binding site prediction [15].

3. Results and Discussion

3.1. In Silico Protein Characterization and Homology Protein Fold Modelling
The DNA sequence of PepYLCV-APWS 723 bp which was translated using BioEdit 7.2.5 software was known to produce 241 amino acids. Similar results were obtained when verified with the NCBI-CDD web tool. ORF AV1 from PepYLCV-APWS shows the conserved domain as the coat protein family / nuclear export factor geminivirus family BR1 from the Pfam acc database. No. Pfam00844 with an E-value of 6.97e-99 [Figure 1]. The results of the analysis also has shown that the 104 N-terminal amino acids of the maize streak virus coat protein bind DNA non-specifically. This family also includes various geminivirus movement proteins that are nuclear export factors or shuttles. One member BR1 facilitates the export of both ds and ss DNA form the nucleus. Similar result also obtained in ToLCV coat protein sequence by Kumar et al. [2012] [16] Papaya Leaf curl virus coat protein by Patel and Kalaria [2018] [17], and ChiLCV coat protein by Mistry et al. [2019] [18].

![Figure 1. Presence of Gemini coat protein as conserved domain in PepYLCV-APWS coat protein](image-url)

Modelling of PepYLCV-APWS coat protein sequences with PDB database result showed 82.11% similarity with coat protein subunit H CryoEM structure of Aqueratum Yellow Vein Virus [6f2s.1 J]. PepYLCV-APWS coat protein sequence was further carried forwarded for homology modelling using swiss model web software for 3D structure prediction [Figure 2.]. The residue visualized in colored ribbon style, start from blue as N-Terminal to red as C-Terminal.
Figure 2. The 3D structure of coat protein PepYLCV-APWS was predicted using Swiss-Model.

The Ramachandran Plot showed 89.7% of residues were in the most favored region, 9.2% in additional allowed regions, 0.0% in generously allowed regions, and 1.0% of residues were in disallowed regions [Figure 3]. The Ramachandran plot visualizes energetically allowed and disallowed regions for the dihedral angles [19]. In poor quality homology models, many residues are found in disallowed regions of the Ramachandran plot. Our data showed only 1.0% of residues were in disallowed regions and so the CP structure was acceptable for further analysis.

Figure 3. Ramachandran Plot shows the quality of the structure. B region represent beta-sheet, A region represents right-handed alpha helix, and L left-handed alpha helix.
### 3.2. Domain and binding site prediction

| No. | SeqID | AA   | Atom | No. | SeqID | AA   | Atom | No. | SeqID | AA   | Atom |
|-----|-------|------|------|-----|-------|------|------|-----|-------|------|------|
| 1   | 24    | TYR  | CD1  | 40  | 89    | ARG  | NH1  | 79  | 184   | VAL  | CG2  |
| 2   | 25    | LYS  | CA   | 41  | 90    | VAL  | CG1  | 80  | 185   | LYS  | CE   |
| 3   | 26    | ARG  | CA   | 42  | 91    | GLY  | CA   | 81  | 186   | ARG  | NH1  |
| 4   | 28    | ALA  | CB   | 43  | 92    | LYS  | CB   | 82  | 187   | PHE  | CB   |
| 5   | 29    | TRP  | N    | 44  | 93    | ARG  | O    | 83  | 188   | PHE  | CB   |
| 6   | 32    | ARG  | CD   | 45  | 94    | PHE  | CZ   | 84  | 193   | TYR  | CE2  |
| 7   | 36    | ARG  | N    | 46  | 96    | VAL  | CG2  | 85  | 194   | VAL  | N    |
| 8   | 37    | LYS  | N    | 47  | 99    | VAL  | O    | 86  | 195   | VAL  | CG2  |
| 9   | 38    | PRO  | CB   | 48  | 100   | TYR  | CE1  | 87  | 197   | ASN  | O    |
| 10  | 40    | LEU  | CB   | 49  | 101   | ILE  | CG1  | 88  | 198   | HIS  | CA   |
| 11  | 41    | TYR  | CE2  | 50  | 108   | ASP  | N    | 89  | 204   | TYR  | CB   |
| 12  | 44    | ARG  | C    | 51  | 110   | ASN  | C    | 90  | 205   | GLU  | CA   |
| 13  | 45    | ARG  | N    | 52  | 111   | ILE  | CG1  | 91  | 207   | HIS  | O    |
| 14  | 46    | THR  | N    | 53  | 114   | LYS  | NZ   | 92  | 209   | GLU  | CG   |
| 15  | 49    | VAL  | CB   | 54  | 118   | ASN  | O    | 93  | 210   | ASN  | CB   |
| 16  | 52    | GLY  | CA   | 55  | 120   | VAL  | CG2  | 94  | 213   | LEU  | CD1  |
| 17  | 53    | CYS  | N    | 56  | 121   | MET  | CE   | 95  | 214   | LEU  | CB   |
| 18  | 55    | GLY  | N    | 57  | 122   | PHE  | O    | 96  | 218   | CYS  | CB   |
| 19  | 56    | PRO  | CG   | 58  | 123   | TRP  | CZ2  | 97  | 220   | HIS  | O    |
| 20  | 57    | CYS  | O    | 59  | 124   | LEU  | CB   | 98  | 221   |ALA   | O    |
| 21  | 58    | LYS  | NZ   | 60  | 127   | ASP  | CB   | 99  | 222   | SER  | C    |
| 22  | 60    | GLN  | CB   | 61  | 128   | ARG  | NE   | 100 | 223   | ASN  | CB   |
| 23  | 62    | PHE  | O    | 62  | 131   | GLY  | CA   | 101 | 224   | PRO  | N    |
| 24  | 64    | GLN  | CB   | 63  | 134   | PRO  | CA   | 102 | 225   | VAL  | CA   |
| 25  | 66    | HIS  | NE2  | 64  | 135   | TYR  | CZ   | 103 | 226   | TYR  | CD1  |
| 26  | 67    | ASP  | CB   | 65  | 140   | LEU  | CD1  | 104 | 229   | LEU  | O    |
| 27  | 69    | THR  | CA   | 66  | 146   | ASN  | CB   | 105 | 231   | ILE  | CG2  |
| 28  | 70    | HIS  | CB   | 67  | 149   | SER  | O    | 106 | 233   | ILE  | CG2  |
| 29  | 71    | THR  | CG2  | 68  | 151   | ALA  | O    | 107 | 235   | PHE  | CD1  |
| 30  | 74    | VAL  | CG1  | 69  | 157   | LEU  | CD2  | 108 | 237   | ASP  | CB   |
| 31  | 75    | LEU  | CA   | 70  | 160   | ARG  | NH2  | 109 | 238   | ASN  | N    |
| 32  | 76    | CYS  | CB   | 71  | 161   | VAL  | CG2  | 110 | 240   | THR  | OG1  |
| 33  | 77    | VAL  | O    | 72  | 165   | HIS  | CE1  |     |       |      |      |
| 34  | 78    | SER  | C    | 73  | 166   | ARG  | NH2  |     |       |      |      |
| 35  | 79    | ASP  | N    | 74  | 167   | PHE  | O    |     |       |      |      |
| 36  | 80    | VAL  | CB   | 75  | 168   | SER  | O    |     |       |      |      |
| 37  | 82    | ARG  | NH1  | 76  | 169   | ALA  | CB   |     |       |      |      |
| 38  | 84    | ASN  | CB   | 77  | 171   | VAL  | CG2  |     |       |      |      |
| 39  | 86    | ILE  | CG2  | 78  | 183   | ILE  | O    |     |       |      |      |

Table 1. The predicted binding key sites of PepYLCV-APWS coat protein structure
In CASTp, voids are defined as buried unfilled space inside proteins after removing all heteroatoms that are inaccessible to water molecules from outside. Pockets are defined as concave caverns with constrictions at the opening on the surface regions of proteins and allow easy access of water probes from the outside. Our data showed the openings were predicted at the start of the N-terminal and the end of the C-terminal [Figure 4.] contains 110 residues in which protein interactions most likely will be involved in these openings area and residues [Table 1.].

**Figure 4.** The surface analysis of coat protein PepYLCV-APWS structure. The mesh showed the openings for pockets and predicted binding site.

4. Conclusion
In this study, surface analysis found predictive binding sites and key residues in the PepYLV-APWS Coat Protein structure. This study aims to predict the 3D structure and protein binding sites of the Alahan Panjang pepper leaf curl virus isolate, West Sumatera [PepYLCV-APWS]. The prediction uses swiss and chimera modeling software for protein modeling. The PepYLCV-APWS coat protein sequences showed 82.11% similarity to the results of the PDB database coat protein subunit H CryoEM structure of *Ageratum Yellow Vein Virus* [6f2s.1 J]. In addition, this research can help to understand the similarities and molecular interactions in order to develop a strong resistance strategy against the gemini virus in the future.

Acknowledgments
This research was conducted with financial support through an Applied Research Scheme research grant with the contract number T / 36 / UN.16.17 / PT.01.03 / AMD / PT Pangan / 2020.

References
[1] Brown, J.K., Fauquet, C.M., Briddon, R.W., Zerbini, M., Moriones, E., and Navas-Castillo, J. 2012. *Family Geminiviridae. In: Virus Taxonomy: Classification and Nomenclature of Viruses—Ninth Report of the International Committee on Taxonomy of Viruses*. King, A.M Elsevier Academic Press.: London.
[2] Fondong, V.N. 2013. Geminivirus protein structure and function. Molecular Plant Pathology, 14 [6]: 635–649.

[3] Gutierrez, C. 1999. Review Geminivirus DNA replication. Genome, 56: 313–329.

[4] Hanley-Bowdoin, L., Bejarano, E.R., Robertson, D., and Mansoor, S. 2013. Geminiviruses: Masters at redirecting and reprogramming plant processes. Nature Reviews Microbiology, 11 [11]: 777–788.

[5] Kumar, R.V. 2019. Plant antiviral immunity against geminiviruses and viral counter-defense for survival. Frontiers in Microbiology, 10 [JUN]: 1–17.

[6] Jeske, H., Lütgemeier, M., and Preiß, W. 2001. DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. EMBO Journal, 20 [21]: 6158–6167.

[7] Pooggin, M.M. 2013. How can plant DNA viruses evade siRNA-directed DNA methylation and silencing? International Journal of Molecular Sciences, 14 [8]: 15233–15259.

[8] Sharma, P., and Ikegami, M. 2010. Tomato leaf curl Java virus V2 protein is a determinant of virulence, hypersensitive response and suppression of posttranscriptional gene silencing. Virology, 396 [1]: 85–93.

[9] Hall TA. 1999. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser., 41[41]: 95–98.

[10] Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. 2020 CDD/SPARCLE: The conserved domain database in 2020. Nucleic Acids Res., 48[D1]: D265–268.

[11] Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. 2017. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. Nucleic Acids Res.:45[D1]: D200–203.

[12] Schwede T, Kopp J, Guex N, Peitsch MC. 2003. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res., 31[13]: 3381–3385

[13] Laskowski, R. A., Watson, J. D. and Thornton, J. M. 2005 ProFunc: A server for predicting protein function from 3D structure. Nucleic Acids Research, 33[SUPPL. 2]: 89–93

[14] Tian, W. et al. 2018. CASTp 3.0: Computed atlas of surface topography of proteins, Nucleic Acids Research. Oxford University Press, 46[W1]: W363–W367.

[15] Pettersen, E. F. et al. 2004. UCSF Chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry, 25[13]: 1605–1612.

[16] Kumar SP, Patel SK, Kapopara RG, Jasrai YT, Pandya HA. Evolutionary and Molecular Aspects of Indian Tomato Leaf Curl Virus Coat Protein, International Journal Plant Genomics, 2012, 1-12.

[17] Patel J, Kalaria RK. In-silico Analysis & Homology of Papaya Ring Spot Virus & Papaya Leaf curl virus coat protein, Bulletin journal of Environment, Pharmacology and Life Science. 2018; 7(8):19-28.

[18] Mistry P, Kalaria RK, Patel AI. Molecular Identification and In silico Characterization of Coat Protein in Chilli Leaf Curl Virus associated in chilli from South Gujarat Region of India. Int J Curr Microbiol Appl Sci. 2019;4(07):711–8.

[19] Gasteiger, E. et al. (2003) ’EXPASy: The proteomics server for in-depth protein knowledge and analysis’, Nucleic Acids Research. doi: 10.1093/nar/gkg563.