Role of the C4 protein PepYLCV in viral symptom determinant: In silico approach

B Nova¹, D Cania², R Hidayati³, T Runifah³, I Chaniago⁴, I Suliansyah⁴, and J Jamsari⁴*

¹ Graduate Program of Agriculture, Faculty of Agriculture, Andalas University
² Undergraduate Program of Agrotechnology, Department of Crop Science, Faculty of Agriculture, Andalas University
³ The Graduate Program of Biotechnology, Multi-discipline graduate Department, Andalas University
⁴ Department of Crop Science, Faculty of Agriculture, Andalas University

Email: jamsari@agr.unand.ac.id

Abstract. The C4 protein of the monopartite begomovirus pepper yellow leaf curl virus [PepYLCV] plays a major role in the pathogenicity of plant virus infection. Complement gene [C4], in virus, is overlapped with complement gene 1 [C1] or known as Replicase gene [Rep]. The C1 protein is directly involved in virus DNA replication, but not the C4 protein. Recent studies suggested that the C4 is a viral symptom determinant but the molecular mechanism has not been elucidated. In this study, the C4 gene of the novel PepYLCV Alahan Panjang was isolated. To further investigate the role of C4 in symptom determinants, in silico approached was used. Computational modeling of C4 was constructed with homology and ab initio modeling software and the model was checked with Ramachandran Plot. The optimum model of the C4 was used in the surface analysis to predict binding sites and key residues. The mutant structure of the C4 protein and its prospect for CRISPR-Cas9 utilization are discussed.

Keyword: PepYLCV, C4, Binding site, Pathogenicity

1. Introduction

Pepper Yellow Leaf Curl Virus [PepYLCV] is a begomovirus that has emerged as one of the largest and devastating plant viruses worldwide [17]. Most begomoviruses infect a wide range of economically important crops like chili, pepper, and tomatoes [6]. PepYLCY is a whitefly-transmitted geminivirus with either two DNA components [bipartite: DNA A and DNA B] or a single DNA component [monopartite] [2,9]. In monopartite, the single-stranded DNA encodes 6 crucial proteins that interact with the host plant both in the early and late infection [C1, C2, C3, C4, V1, and V2] [20].

The C4 protein in monopartite begomoviruses has important functions. Previous studies have indicated that C4 is a major viral determinant [14]. The C4 has been shown to interact with the receptor-like kinase the receptor-like kinases BARELY ANY MERISTEM 1 [BAM1] and greatly impact RNA interference [RNAi] antiviral defense mechanism [13,17]. BAM 1, with its homolog BAM 2, is responsible for systemically cell-to-cell spread of the RNA interference response via plasmodesmata and suppress virus infection. Begomovirus C4 is found in plasmodesmata [PD] and is known to inhibit certain functions of BAM 1 associated with intercellular RNAi spread [19].
2. Materials and Method

2.1. DNA and Protein Sequence
The PepYLCV C4 DNA sequence was obtained from Plant Virus Genome Collection, Laboratory of Biotechnology, Andalas University. The receptor-like kinase NIK1 DNA sequence was obtained from the NCBI database. DNA to protein sequence was translated with ExPASy [7].

2.2. In Silico Ab-Initia And Homology Protein Fold Modelling
The three-dimensional structure of PepYCLV-BC1 was predicted using Phyre2 [10]. Quality assessment of the predicted structure was performed with ProFunc to generate The Ramachandran plot [11].

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 [21].

2.4. Docking Protein-Protein Interaction
PatchDock was used to predict protein-protein interaction between the C4 and BAM 1 predicted structure [4].

2.5. In Silico Protein Visualization
Chimera 1.14 was performed to visualize protein modeling, binding site prediction, and protein-protein docking [15].

3. Results and Discussion

3.1. In Silico Ab-Initia And Homology Protein Fold Modelling
One reference protein, Oxidoreductase [PDB ID 1W7C] was used to model the structure of C4 protein. The identity and confidence scores comparing to the reference were 16 % and 41.5%, respectively. A low level of sequence identity indicates less accurate alignment between the reference and target sequence. Only 25 of the 84 target amino acids were successfully aligned. The structure homologs cannot be identified completely because the identity score was low, so the structure has to be modeled without any references called ab initio modeling. The remaining 59 amino acids were modeled ab initio which results in a novel fold.

In ab initio modeling, the Ca atom, the Cβ atom, the peptide bond, or at the center of mass of the side chains were assessed for its potential interactions [8]. The interactions were used as a reference for creating conformation space and predicting the most accurate predicted structure [5]. The ab initio modeling, despite generating a low-resolution structure, still sufficiently able to inform about protein function via analysis of local sequence or structure patterns [3]. We concluded that the confidence score of 41.5%, was in an acceptable range, thus the structure could be used for the analysis. In Figure 1. [a,b] the structure of C4 has one helix, one beta-hairpin, two strands, twenty-two beta turns, and 1 gamma turn.
Figure 1. The Ab Initio and Homology Fold Protein Modelling of C4 Protein. [a] The secondary structure of the C4 structure. [b] The residue is visualized in colored ribbon style, start from blue as N-Terminal to red as C-Terminal. [c] Ramachandran Plot shows the quality of the structure. B region represents beta-sheet, A region represents right-handed alpha-helix, and L left-handed alpha helix.

The Ramachandran Plot showed 65.3% of residues were in the most favored region, 22.2% in additional allowed regions, 11.1% in generously allowed regions, and 1.4% of residues were in disallowed regions [Figure 1. [c]]. The Ramachandran plot visualizes energetically allowed and disallowed regions for the dihedral angles [16]. In poor quality homology models, many residues are found in disallowed regions of the Ramachandran plot. Our data showed only 1.4% of residues were in disallowed regions and so the C4 structure was acceptable for further analysis.

3.2. Domain and binding site prediction

The pocket analysis shows the C4 structure has 7 openings, with pocket and mouth MS area score 1059.5 and 276.2, respectively.
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 [1,12]. Our data showed the openings were predicted at the start of the N-terminal and the end of the C-terminal [Figure 2, (a)] contains 37 residues in which protein interactions most likely will be involved in these openings area and residues [Table 1]. The N terminal of C4 is essential for PD localization, specifically the myristoylation motif. The alteration of myristoylation consensus [MGCXXSK/T], especially G>A, in the N terminal abolished infectivity in tomato [17]. The structural models of the C4 mutant [G2A] were generated by homology and ab initio modeling. The single mutation of Glycine to Alanine in myristoylation consensus generating a different structure compared to the wild type. Also, the surface analysis showed significantly different sites [Figure 2, (b)] and contains only 18 residues [Table 1]. The mutation substantially affects the secondary structure and gives rise to major disruption of the C4 Wild type structure and predicted binding sites. Substitution of the important residue by mutating Gly-2 to Ala-2 leads to the tremendous change of the predicted binding site and completely diminished the number of residues involved in potential protein interactions.

Table 1. The predicted binding key sites of C4 Wild type and Mutant [G2A] structure

| SeqID | AA | Atom | SeqID | AA | Atom | SeqID | AA | Atom |
|-------|----|------|-------|----|------|-------|----|------|
| 1     | MET| O    | 3     | LEU| O    | 5     | ILE| CA   |
| 2     | GLY| CA   | 4     | LEU| O    | 6     | SER| N    |
| 3     | LEU| N    | 7     | THR| CA   | 8     | CYS| N    |
| 4     | LEU| N    | 9     | LEU| N    | 11    | ASN| CB   |
| 6     | SER| N    | 12    |     |      | 18    | ALA| O    |
| 7     | THR| N    | 13    |     |      | 19    | ARG| CA   |

Figure 2. The surface analysis of [a] C4 Wild type and [b] Mutant [G2A] structure. The mesh showed the openings for pockets and predicted binding site.
4. Conclusion
In this study, the surfacing analysis found predicted binding sites and key residues in the PepYLV C4 protein structure. We found that Gly-2 to Ala-2 mutation has a high potential to affect protein interaction and the symptom development of virus infection. This single mutation is the potential to interrupt the function of C4 protein and inhibit the mobility of viruses.

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] Binkowski, T. A., Naghibzadeh, S. and Liang, J. [2003] ‘CASTp: Computed Atlas of Surface Topography of proteins’, *Nucleic Acids Research*.
[2] Bisaro, D. M. *et al.* [1982] ‘Molecular cloning and characterisation of the two DNA components of tomato golden mosaic virus’, *Nucleic Acids Research*.
[3] Bonneau, R. *et al.* [2001] ‘Functional inferences from blind ab initio protein structure predictions’, *Journal of Structural Biology*. 
[4] Duhovny, D., Nussinov, R. and Wolfson, H. J. [2002] ‘Efficient Unbound Docking of Rigid Molecules’, Lecture Notes in Computer Science, 2452, pp. 185–200.

[5] Eastwood, M. P. et al. [2001] ‘Evaluating protein structure-prediction schemes using energy landscape theory’, IBM Journal of Research and Development.

[6] Garcia-Arenal, F. and Zerbini, F. M. [2019] ‘Life on the Edge: Geminiviruses at the Interface between Crops and Wild Plant Hosts’, Annual Review of Virology.

[7] Gasteiger, E. et al. [2003] ‘ExPASy: The proteomics server for in-depth protein knowledge and analysis’, Nucleic Acids Research.

[8] Hardin, C., Pogorelov, T. V. and Luthey-Schulten, Z. [2002] ‘Ab initio protein structure prediction’, Current Opinion in Structural Biology.

[9] Harrison, B. D. and Robinson, D. J. [1999] ‘Natural genomic and antigenic variation in whitefly-transmitted geminiviruses [Begomoviruses]’, Annual Review of Phytopathology.

[10] Kelley, L. A. et al. [2015] ‘The Phyre2 web portal for protein modeling, prediction and analysis’, Nature Protocols. Nature Publishing Group, 10[6], pp. 845–858.

[11] 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], pp. 89–93.

[12] Lee, B. and Richards, F. M. [1971] ‘The interpretation of protein structures: Estimation of static accessibility’, Journal of Molecular Biology.

[13] Li, Z. et al. [2020] ‘C4, the Pathogenic Determinant of Tomato Leaf Curl Guangdong Virus, May Suppress Post-transcriptional Gene Silencing by Interacting With BAM1 Protein’, Frontiers in Microbiology.

[14] Mei, Y. et al. [2018] ‘Nucleocytoplasmic Shuttling of Geminivirus C4 Protein Mediated by Phosphorylation and Myristoylation Is Critical for Viral Pathogenicity’, Molecular Plant.

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

[16] Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. [1963] ‘Stereochemistry of polypeptide chain configurations’, Journal of Molecular Biology.

[17] Rojas, M. R. et al. [2018] ‘World Management of Geminiviruses’, Annual Review of Phytopathology.

[18] Rosas-Diaz, T. et al. [2017] ‘A plant receptor-like kinase promotes cell-to-cell spread of RNAi and is targeted by a virus’, bioRxiv.

[19] Rosas-Diaz, T. et al. [2018] ‘A virus-targeted plant receptor-like kinase promotes cell-to-cell spread of RNAi’, Proceedings of the National Academy of Sciences of the United States of America.

[20] Ruhel, R. and Chakraborty, S. [2019] ‘Multifunctional roles of geminivirus encoded replication initiator protein’, VirusDisease.

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