Simulated mutations in the geminivirus replicase gene using in-silico CRISPR / Cas9-based methods

R Hidayati¹, B Nova², D Cania³, T Runifah¹, D H Tjong¹, and J Jamsari³∗

1  The Graduate Program of Biotechnology, Multi-discipline graduate Department, Andalas University

2  Graduate Program of Agriculture, Faculty of Agriculture, Andalas University

3  Undergraduate Program of Agrotechnology, Department of Crop Science, Faculty of Agriculture, Andalas University

4  Department of Crop Science, Faculty of Agriculture, Andalas University

Email: jamsari@agr.unand.ac.id

Abstract. Geminivirus is one of the causes of losses in the agricultural sector. This virus infects many horticultural crops such as tomatoes, potatoes, eggplant, cucumbers, and chilies. All genes in geminivirus have an important and interrelated role in the success of infection. Replicase is a gene that is key in the replication of the viral genome in its host. The occurrence of mutations in this gene is thought to stop the virus from multiplying in the host. This will lead to the failure of the virus to infect the plant. To achieve this, an in-silico mutation simulation can be performed first. Mutations were simulated following the CRISPR / Cas9 method using Phyre2, ProFunc, CASTp, and Chimera software. This simulation succeeded in predicting that the CRISPR / Cas9 system on the replicase gene could inhibit viral replication. The success of this simulation will become a reference and consideration for CRISPR/Cas9 further research at the laboratory scale.

Keyword: Chimera, mutation, replicase.

1. Introduction

Geminivirus is a single strand virus that replicates through an intermediary in the nucleus of a plant cell[1]. Due to their small genome, the geminivirus only provides the factor necessary to initiate replication rolling circles and relies on plant nucleus DNA polymerase to amplify its genome [2]. Rep [also called AC1 or C1] is the only viral protein essential for viral replication [3]. Rep binds to specific DNA sequences, catalyzes DNA division, and ligates to initiate and stop rolling circle replication. Rep is often involved in protein-protein interactions [4]. Rep also interacts with host factors involved in cell division and differentiation [5]. pRBR is one of the substrates of the cyclin-dependent kinase and is high in differentiated tissue [6] and RBS plays a role in being responsible for ribosome recruitment during translation initiation. Rep binds to mitosis and protein kinases associated with early leaf development, but the role of these interactions during infection has not been seen. Conversely, there is evidence that binding of Rep to pRBR alters host transcription control to induce replication of plant DNA synthesis [6]. The interaction of Rep and RBS can change KEE 146 amino acids. This needs further attention to see the effect of these changes on protein Rep.
2. Materials and Method

2.1. Protein Sequence Collecting.
In this study, the virus was isolated from chili plants infected by geminivirus in Alahan Panjang [PepYLCV-APWS]. The genetic data of Rep protein was chosen from sequence data of Geminivirus PepYLCV-APWS. All sequence data were used in fasta format.

2.2. 2D Structure Prediction
The 2D structure prediction was performed to determine the estimated conformation protein structure, helix, and sheet patterns. The prediction generated using the ProFunc software from EMBL-EBI [http://www.ebi.ac.uk/thorntonrv/databases/cgi-bin/profunc/] [7].

2.3. Protein Modeling and Binding Site Prediction
The 3D protein prediction of the structure was built using the Phyre2 software [8]. Then for surface analysis is performed with CASTp 3.0 to predicts pockets and caves in predicted proteins and determines the prediction of binding sites [9].

2.4. Visualization of Protein Modelling and Protein Mutation
The UCSF Chimera application was performed for visualization of protein predictions, prediction of binding sites, and protein mutation [10].

3. Results and Discussion

3.1. 2D Structure Prediction Protein
Predict the 2D protein structure using ProFunc software by uploading a .pdb file to the site. The Rep protein predicted 2D structure to determine the helix and sheet pattern of the protein. This pattern determines the conformation and indentation of the protein.

![Figure 1. Comparison of Rep protein 2D structure: [A] wild type and [B] mutants.](image)
Figure 1 shows the comparison of 2D structure between Rep wild type and mutant shows that it changes the pattern and structure of the protein. There is more α helix in the mutant Rep and the number of β sheets is the same in Rep wild type and mutant Rep. The number of α helix and β sheets in the protein determines the pockets and caves of the protein. The difference in the number of helices and sheets causes changes in the binding capacity of proteins to viral molecules. This is thought to cause no interaction between the Rep protein and the virus.

3.2. Modelling Protein
Protein modeling was conducted to see the prediction of protein form. Files with the .pdb format are opened using the UCSF Chimera application for visualization.

![Figure 2. Comparison 3D protein structure of (A) Rep wild type, and (B) Rep mutant. The residue is visualized in colored ribbon style, start from blue as N-Terminal to red as C-Terminal.](image)

Figure 2 shows the comparison of the wild-type Rep protein [Figure 2A] and mutant [Figure 2B] predicted by 3D structure using Phyre2 software. Rep wild type [PDB ID: IL51] and mutants [PDB ID: 6Q1M] have 340 residues, 99% of which are successfully modeled with > 90% accuracy. This figure only shows the difference in the prediction of the 3D protein structure. There are differences in the conformation and shape of these proteins due to the difference in the number of helix and sheets.

3.3. Binding site prediction
The mutation simulation in this study is the change of the amino acid Glutamine to Leucine. Binding site prediction is done by uploading the .pdb file format in the CASTp software. This prediction is done to determine the position and score of the interaction area based on surface analysis[11]. The pocket analysis shows the Rep protein wild type and mutant had pockets with the highest MS mouth area scores, 45.8 and 930.9 respectively.
Figure 3. The surface analysis of [A]Rep Wild type and [B] Mutant [KEE$_{146}$] structure. The mesh showed the openings for pockets and predicted binding site.

The geometric and topological properties of the protein structure, including surface pockets, interior cavities, and cross channels, are very important for proteins to carry out their functions. Computed Atlas of Surface Topography of protein [CASTp] helps in finding, describing, and measuring the geometric and topological properties of protein structures[12]. This software provides reliable and comprehensive identification and quantification of protein topography. In Figure 3 it is shown that the mutant Rep has less surface interaction than the wild type Rep. This change is thought to hinder the interaction of Rep with important molecules related to regulation. Based on the binding site prediction, it is known that the mutant Rep that converts Glu to Leu is quite effective in reducing protein interactions. It can be seen that there is a decrease in the predicted binding area. This of course narrows down the possibility of Rep interactions with viral molecules. To ensure this, it is necessary to carry out directional mutations in vivo.

4. Conclusion
In this study, the surfacing analysis found predicted binding sites and key residues in the PepYLCSV C1 protein structure. Mutations that convert Glu to Leu are effective in reducing protein interactions by decreasing the predicted binding area. This single mutation potentially interferes with the replication of the C1 gene in plants thus inhibiting viral replication in plant cells.

Acknowledgment
This research was funded by an Applied Research Grant with contract number T / 36 / UN.16.17 / PT.01.03./AMD/ PT Pangan / 2020.

References
[1] Briddon R W 2015 Geminiviridae eLS pp 1–12
[2] Krupovic M, Ravantti J J and Bamford D H 2009 Geminiviruses: a tale of a plasmid becoming a virus. *BMC Evol. Biol.* 9 112
[3] Baltes N J, Gil-Humanes J, Cermak T, Atkins P a and Voytas D F 2014 DNA Replicons for Plant Genome Engineering *Plant Cell* 26 151–63
[4] Gutierrez C 2000 DNA replication and cell cycle in plants: Learning from geminiviruses *EMBO J.*
[5] Nordman J and Orr-Weaver T L 2012 Regulation of DNA replication during development
Development

[6] Arguello-Astorga G, Lopez-Ochoa L, Kong L J, Orozco B M, Settlage S B and Hanley-Bowdoin L 2004 A novel motif in geminivirus replication proteins interacts with the plant retinoblastoma-related protein J. Virol. 78 4817–26

[7] Laskowski R A, Watson J D and Thornton J M 2005 ProFunc: A server for predicting protein function from 3D structure Nucleic Acids Res.

[8] Kelley L A, Mezulis S, Yates C M, Wass M N and Sternberg M J E 2015 The Phyre2 web portal for protein modeling, prediction and analysis Nat. Protoc. 10 845–58

[9] Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y and Liang J 2006 CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues Nucleic Acids Res. 34

[10] Pettersen E F, Goddard T D, Huang C C, Couch G S, Greenblatt D M, Meng E C and Ferrin T E 2004 UCSF Chimera — A Visualization System for Exploratory Research and Analysis

[11] Mathew A, Verma A and Gaur S 2014 An in-silico insight into the characteristics of β-propeller phytase Interdiscip. Sci. Comput. Life Sci. 6 133–9

[12] Tian W, Chen C, Lei X, Zhao J and Liang J 2018 CASTp 3.0: Computed atlas of surface topography of proteins Nucleic Acids Res. 46 W363–7