Structural Modeling and Validation of Rep protein of Begomovirus Strains (TLCBV and TYLCTHV)

Rajneesh Prajapat, Vivek Bajpai and Rajarshi Kumar Gaur*

Department of Science, FASC, Mody Institute of Technology and Science, Lakshmangarh, Sikar - 332311, Rajasthan (India).

* Corresponding author e-mail: gaurrajarshi@hotmail.com

ABSTRACT

Homology modeling involves taking a known sequence with an unknown structure and mapping it against a known structure of one or several similar (homologous) proteins. It would be expected that two proteins of similar origin and function would have reasonable structural similarity. Therefore it is possible to use the known structure as a template for modeling the structure of the unknown structure. Proteins those share same function generally have similar structures. During alignment if two proteins show maximum sequence identity they also show similar folding pattern. This principal is became the foundation of homology modeling.
The Geminivirus taxonomic group of plant viruses is characterized by geminate particles and genomes consisting of single-stranded circular DNA molecules of about 2.5 to 2.8 kb in size. Agricultural plants are threatened by many diseases caused by whitefly-transmitted geminiviruses. Since these diseases are in a fast spreading phase, it is urgent to device rapid diagnosis methods and to produce resistant plants.

We have analyzed FASTA protein sequences of Rep protein of Begomovirus strains including Tomato leaf curl Bangalore virus (TLCBV) and Tomato yellow leaf curl Thailand virus (TYLCTHV), a whitefly-transmitted geminivirus. We find out the position of tyrosine with the help of Chimera, explain that there are difference in tyrosine position in Rep proteins of TLCBV and TYLCTHV because tyrosine play important role in DNA replication. Phylogram (Constructed using clustal-w) and multiple sequence alignment results of Q6T874 and Q9WPF9 suggest that these two sequences have evolutionary relationship. Homology modeling principles were used to predict the three dimensional structure, stability of replication associated protein of both TLCBV and TYLCTHV.

Keywords: Geminivirus, Begomovirus, TLCBV, TYLCTHV, PDB, Homology modeling

INTRODUCTION

By Experiment we can predict protein structure, analysis of protein function, interaction, antigenic behavior and rational design of protein with increased stability. But unlike experimental structure, protein modeling are use to predict a structure from its sequence with an accuracy that allow users to use safely generated *in silico* protein models. Several proteins are very large for NMR analysis and cannot be crystallized for X-ray diffraction.
Therefore protein modeling is the only way to obtain structural information if experimental techniques fail.
Geminiviruses are single - stranded DNA (ssDNA) viruses that cause severe disease in major crop plants worldwide. Most of *geminiviruses* belong to the genus *Begomovirus*, which are transmitted exclusively by the whitefly *Bemisia tabaci* (Harrison and Robinson, 1999).
By using the homology modeling principles we can find the 3D structure of Rep protein of TLCBV and TYLCTHV and also predict the stability of the protein.

MATERIALS AND METHODS

Different homology modeling methods have different approaches to construct 3D model for given templates and alignments.
Database searching required for retrieve of protein (Replication associated protein of TYLCNDV and TYLCTHV) which lack PDB entry or in other words sequence which do not have experimentally determined structure. The sequence retrieved must be aligned with the protein of known structure present in Protein Data Base (PDB). The methods used for homology modeling can be broadly divided into four components:

1. Database search
2. Sequence analysis and alignment
3. Homology modeling search
4. Model evaluation and analysis

Toole Used:

ExPASy (Expert Protein Analysis System), NCBI (National Center for Biotechnology Information), PDB (Protein Data Bank), 3D-JIGSAWN (Protein Comparative Modeling Server), Verify 3D Structure Evaluation server, UCLA-DOE Server , SAVA (Structure Analysis and Validation server), CLUSTAL – W and CHIMERA.

Fig. 1. FASTA sequence of Rep protein (Q6T874) of Tomato leaf curl Bangalore virus (TLCBV).

>tr|Q6T874|Q6T874_9GEMI Rep protein OS=**Tomato leaf curl Bangalore virus** GN=**AC1** PE=4 SV=1
MPAPNRFKINAKNYFLTYPKCSLTKEEALSQLLNLQTPTSKKFIRICRELHEDGTPHLHVLIQFEGKFQCKNNRF
FDLTSPTRSAHFHPNIQGAKSASDVKSYMEKDGDVIDHGVFQVDGRSARGGCQTANDAYAEAINSGSKAEAL
NILKEKAPRDFLLQFHNLNSNLDRYFQEPPAPYVSPFHLSSFDTVPEELYQWAAENVVDEAAR
PIRPISIVIEGDSRTGKTMWARSLGLHNYLCGHLDLSPRVYSNDAWYNVIDDVDPHYLKHFKEFMGAQRDWQ
SNTKYGKPVQIKGGIPTIFLCNPGPNSSYKEFLDEDKNNALKQWTLKNARFITLEEPLYSGSNKSATQPSQEGD
QASTS

Fig. 2. FASTA sequence of Rep protein (Q9WPF9) of Tomato yellow leaf curl Thailand virus (TYLCTHV).

>tr|Q9WPF9|Q9WPF9_9GEMI Rep protein OS= **Tomato yellow leaf curl Thailand virus**-[2] GN=**AC1** PE=4 SV=1
MPPSKKFLINAKNYFLTYPHCSLTKEEEALSQILNLSTPTNKLIIRICRELHEDGTPHLHLLIQFEGKFKCQNNRF
FDLTSPTRSAHFHPNIQGAKSSTDVKAYMEKDGDVLDHGIFQIDGRSARGGCQSANDAYAEAINSGSKASALN
ILREKAPKDFVLQFHNLNSNLDRIFTPPIEEYISPFSSSSFNQVPEELEEWACNNVLSRARPLRPITIVIEGDSRTG
KTMWARSLGPHNYLCGHLDLSPKVYNNDAWFNVIDDVDPHYLKHFKEFMGAQRDWQSNTKYGKPVQIKG
GIPTIFLCNPGPNSSYKEYLEEEKNSALRNWAIKNAIFVTLEGPLYSGTYQGATPNSQEDNQTTES

Table 1: Positions of Tyrosine (TYR) in Replication-associated protein of Tomato leaf curl Bangalore virus (TLCBV) and Tomato yellow leaf curl Thailand virus (TYLCTHV).

S.No	Protein and Strain	Gene	Accession Number	Position of Tyrosine in .PDB file of Protein
1.	Rep protein (Q6T874) of TLCBV	AC1	Q6T874	72, 76, 161, 192
2.	Rep protein (Q9WPF9) of TYLCTHV	AC1	Q9WPF9	81, 85, 171, 202

Fig. 3. Position of tyrosine in Rep protein (Q6T874) of TLCBV (created by CHIMERA).

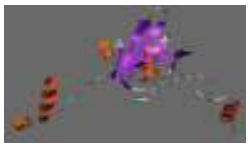

Fig. 4. Position of tyrosine in Rep protein (Q9WPF9) of TYLCTHV (created by CHIMERA).

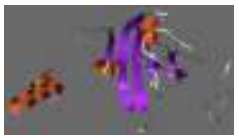

Fig. 5. Profile search plot of Rep protein of TLCBV (81.14% of the residues had an averaged 3D-1D score > 0.2) (http://nihserver.mbi.ucla.edu/SAVES_3/saves.php)

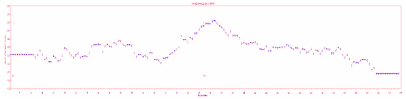

Fig.6. Profile search plot of Rep protein of TYLCTHV (59.88% of the residues had an averaged 3D-1D score > 0.2) The Profile search plot of both of the proteins shows the comparative stability (http://nihserver.mbi.ucla.edu/SAVES_3/saves.php)

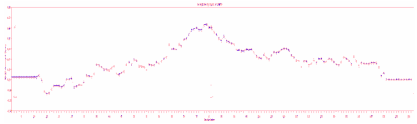

Fig. 7. Ramanchandran plot of Rep protein (Q6T874) of TLCBV. Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favouredregions. (http://nihserver.mbi.ucla.edu/SAVES_3/jobs/1565717/procheck/Q6T874_01.pdf)

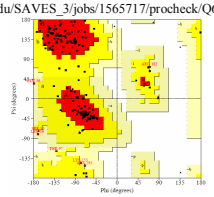

Fig. 8. Ramanchandran plot of Rep protein (Q9WPF9) of TYLCTHV. Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favouredregions. (http://nihserver.mbi.ucla.edu/SAVES_3/jobs/506550/procheck/Q9WPF9_01.pdf)

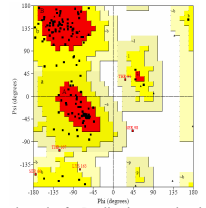

Table 2: Results of Ramachandran plot for Replication associated protein of TYLCNDV and TYLCTHV.

S. No	Protein	Core (%)	Allowed (%)	Gener (%)	Disallowed (%)
1	Rep protein (Q6T874) of TLCBV	78.7	16.7	4.7	0.0
2	Rep protein (Q9WPF9) of TYLCTHV	72.1	24.3	2.9	0.7

CONCLUSION

In this paper, we described application of homology modeling, because it is an effective way to obtain useful information about the proteins of interest. Structure of unknown proteins can be identified on the basis of amino acid sequence pattern matching of both known and unknown proteins. By selecting best homologous proteins showing maximum sequence similarity with the unknown protein, we may predict the structure of unknown protein and with the help of Profile search and Ramachandran plot we can check the stability of proteins.
Phylogram of FASTA sequence of Q6T874 and Q9WPF9 (361 amino acid in both) explain the evolutionary relationships among sequences. The Profile search plot of both of the proteins shows the comparative stability. Ramachandran plots of Rep protein (Q6T874) of TLCBV have 78.7 % amino acid in core region, 16.7 % in allowed, 4.7 % in gener and 0.0 % in disallowed region. Ramachandran plots of Rep protein (Q9WPF9) of TYLCTHV have 72.1% amino acid in core region, 24.3% in allowed, 2.9% in gener and 0.7% in disallowed region; these comparative results explain that Rep protein (Q6T874) of TLCBV more stable than Rep protein (Q9WPF9) of TYLCTHV. Profile search plots of Q6T874 and Q9WPF9 explain that these two proteins are energetically stable.

ACKNOWELDGMENT

RKG is thankful to DBT, New Delhi for providing the financial support during this study. The authors are also thankful to Prof. Shakti Baijal for her valuable suggestion.

REFERENCES

1. Arguello-Astorga, G.R., Guevara-Gonzalez, R.G., Herrera-Estrella, L.R. 1994. ,Rivera-Bustamante, R.F. Geminivirus replication origins have a group-specific organization of iterative elements – a model for replication. *Virology*, 203:90–100.
2. Chatterji, A., Beachy, R.N., Fauquet, C.M. 2001. Expression of the oligomerization domain of the replication-associated protein (Rep) of tomato leaf curl New Delhi virus interferes with DNA accumulation of heterologous geminiviruses. *J. Biol. Chem*, 276:25631-25638.
3. Prajapat R., Gaur, R.K., Raizada, R., Gupta, V.K.2010. *In silico* Analysis of Genetic Diversity of Begomovirus using Homology Modelling. *Journal of Biological Sciences*, 10 (3): 217-223.