Research Article

Evolutionary and Molecular Aspects of Indian Tomato Leaf Curl Virus Coat Protein

Sivakumar Prasanth Kumar,1 Saumya K. Patel,1 Ravi G. Kapopara,1 Yogesh T. Jasrai,1,2 and Himanshu A. Pandya1,2

1 Department of Bioinformatics, Applied Botany Center, University School of Sciences, Gujarat University, Ahmedabad 380 009, India
2 Department of Botany, University School of Sciences, Gujarat University, Ahmedabad 380 009, India

Correspondence should be addressed to Yogesh T. Jasrai, yjasrai@yahoo.com

Received 24 August 2012; Revised 7 November 2012; Accepted 8 November 2012

1. Introduction

Tomato leaf curl virus (ToLCV) is one of the most devastating causal agents of tomato (Solanum lycopersicum) crop which had emerged causing damage and encroaching new areas in tropical and subtropical continents every year. Plant-infecting geminiviruses belong to the family Geminiviridae in which Begomovirus is one among the genera possessing both mono- and bipartite genomes that infect especially dicotyledonous plant species [1]. The disease is marked by symptoms such as yellowing of leaf lamina with upward leaf curl as well as distortion, reduction in internodes, new leaves size reduction, wrinkle facade, stunted growth, and dissemination of flower from plant before onset of fruiting. ToLCV is primarily transmitted by sweet potato whitefly (Bemisia tabaci) and silver leaf whitefly (also called Biotype B; Bemisia argentifolii). Whiteflies harboring virus can nonspecifically infect a wide spectrum of plant crops and weeds including eggplant, potato, tobacco, pepper, and common bean. Infected plants seem healthy but develop symptoms leading to enormous economic loss [2].

In Indian subcontinent, ToLCV is a major problem for tomato-growing regions as several reports on new strains have been documented including New Delhi, Lucknow, Bangalore, Varanasi, Mirzapur, Vadodara, and so forth and posed a threat to crop productivity [6]. Indian ToLCV isolates are mostly monopartite (DNA-A) in nature with few isolates possessing bipartite (DNA-A and DNA-B) genome organization such as tomato leaf curl New Delhi virus (ToLCNDV) and tomato leaf curl Palampur virus (ToLCPalV) [7]. Both DNA-A and DNA-B are single-stranded (ss) DNA genomes of approximately 2.7 kb size and encode viral factors essential for viral replication, encapsidation, transmission, and systemic spread [8]. Jyothsna et al. 2012 reported tomato leaf curl Gujarat virus (ToLCGV) possesses monopartite genome and is infectious expressing systemic symptoms in Nicotiana benthamiana and tomato [9]. An increased symptom severity and shortened incubation period required for symptom
expression was noticed when ToLCGV was coinoculated with betasatellite of tomato yellow leaf curl virus Thailand (TYLCTHB) resulted in yellow mottling [9]. The molecular relationship of ToLCGV-[Vad], an isolate from Vadodara district of Gujarat, with other strains revealed that it belongs to Old World Begomoviruses and established a closely related cluster with other North Indian strains including ToLCGV-(Varanasi)-[Var] and ToLCGV-(Mirzapur)-[Mir] based on the DNA-A sequence alignment [10].

The measure “breeding for resistance” conceptualizes the introduction of resistance genes found in wild tomato species into tomato cultivars to develop resistance against diseases. Kunik et al. 1994 demonstrated that tomato plants transformed with TYLCV coat protein were found to be virus-resistant [11]. In India, Agrobacterium tumefaciens mediated transformation of coat protein gene was carried out to develop ToLCV tolerant/resistant transgenic tomato plants under glass house conditions [12]. Transgenic tomato plants containing cucumber mosaic virus coat protein gene was also successfully transformed [13]. An asymmetric synergism and virulent pseudorecombinant between ToLCNDV and ToLCGV was reported by Chakraborty et al. 2008 and found enhanced pathogenicity when tested in N. benthamiana, N. tabacum, and S. lycopersicum [14]. An evidence for natural recombination was observed between tomato leaf curl Bangalore virus (ToLCBV), ToLCBV [Ban 5], and ToLCBV [Kolar] and examined the possibility of recombination between strains/species that coexist within the same geographical location [15]. Hence, tremendous consideration should be given to study the biological and molecular properties of this newly emerging causal agent.

In the present study, we examined the evolutionary and molecular prospects of ToLCGV-[Vad] coat protein. Sequence analysis of coat protein revealed that amino acids essential for systemic infection, viral particle formation, and insect transmission to host cells were evolutionarily compatible when compared to non-Indian isolates giving clues of evolutionary conservativeness. Further, molecular modeling of coat protein provided a topology similar to characteristic Geminate viral particle. Electronic properties of coat protein facilitated its interaction with viral DNA with the loop element acting as molecular recognition interface which is facilitated its interaction with viral DNA with the loop Emboss stretcher (http://www.ebi.ac.uk/Tools/psa/emboss_stretcher/) [22] with a representative family protein member (BR1 nuclear shuttle protein from squash leaf curl virus [SQLCV]; NCBI accession No. NP_047247.2) against the coat protein of study. Both sequences were aligned using EMBLOSUM62 scoring matrix with a gap opening and extending penalty of 12 and 2.

2. Materials and Methods

2.1. Protein Sequence Retrieval. The coat protein of ToLCGV-[Vad] (accession no. AAL78666.1) was retrieved from NCBI database (http://www.ncbi.nlm.nih.gov/) [16]. Coat protein sequences of Indian strains (Bangalore-CAA88227.1, Bangalore (Ban4)-AAD51286.1, Bangalore (Ban5)-AAK19178.1, Bangalore (Kolar)-AAL26553.1, Varanasi-AAO25668.1, Kelloo-AAM21566.1, Karnataka-AAO89293.1, New Delhi (Mild)-AAA92817.1, and Lucknow-CAA76209.1) were also obtained for multiple sequence alignment and phylogenetic analysis.

2.2. Protein Family Classification. The family of coat protein was studied using a combination of programs, namely, NCBI CD-Search (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi, database searched: CDD v3.03–42251 PSSMs) [17], PSI-BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) [18], and Pfam (http://pfam.sanger.ac.uk/) [19]. CD-Search is a NCBI’s interface to search Conserved Domain Database (CDD) which utilizes RPS-BLAST (Reverse-PSI-BLAST; a variant of PSI-BLAST) to scan a set of precalculated position specific scoring matrices (PSSMs) using a protein query. PSI-BLAST (position-specific-iterated BLAST) uses initial matches to query sequence to build scoring matrix and appends additional matches to the matrix by an iterative search method in order to detect remote homologs. Pfam designates protein family by HMM (Hidden Markov Model)-based search (default settings were chosen and Pfam-A significant matches were only considered) over known protein family classifiers.

2.3. Analysis of Nuclear Localization Signal and Its Prediction. Nuclear localization signals (NLSs) were predicted using cNLS Mapper (http://nls-mapper.iab.keio.ac.jp/) [20] as coat protein that is known to be karyophilic [21]. cNLS Mapper is a computer program that predicts NLS by activity-based profile search and an additivity-based motif scoring function in different classes of importin-α/β pathway-specific NLS. The prediction was made with a score cut-off of 5.0 and searched for both mono- and bipartite NLSs with a long linker (13–20 amino acid length) as ToLCGV possesses mono-bipartite genome organization [6]. Classic NLS typically rich in basic amino acids such as lysine and arginine, the counts of basic amino acids was performed manually in the above predicted NLSs. Comparison with literature-reported NLS specific to BR1 nuclear export family was carried out to examine the pattern of nuclear localization. This was achieved using pairwise sequence alignment using EMBL Stretcher (http://www.ebi.ac.uk/Tools/psa/emboss_stretcher/) with a representative family protein member (BR1 nuclear shuttle protein from squash leaf curl virus [SQLCV]; NCBI accession No. NP_047247.2) against the coat protein of study. Both sequences were aligned using EMBLOSUM62 scoring matrix with a gap opening and extending penalty of 12 and 2.

2.4. Multiple Sequence Alignment and Phylogenetic Analysis of Coat Proteins. Coat protein sequences of Indian strains were used for the analysis of multiple sequence alignment (MSA). MSA was performed using EBI ClustalW program (http://www.ebi.ac.uk/Tools/msa/clustalw2/) [23] in which the sequences were aligned pairwise initially (gap open penalty = 10, gap extension penalty = 0.1, matrix = Gonnet) and then the best local pairs (gap open penalty = 10, gap extension penalty = 0.20, matrix = Gonnet) were clustered by Neighbour-joining (NJ) technique. Subsequently, an alignment file in ClustalW format was generated and specified as input to draw phylogenetic tree using Phylip version 3.68 package (http://evolution.genetics.washington.edu/phylip/) [24]. NJ algorithm was used to draw tree with inclusion of branch length.
2.5. Structure Modeling of Coat Protein. Sequence-based similarity searching was initially executed in NCBI non-redundant (NR) database using BLASTp program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) [25] with default settings to find a close homolog with known 3D protein structure information is known. Similarly, BLAST based homolog search in RCSB Protein Data Bank (PDB; http://www.rcsb.org/pdb/home/home.do) [26] was also carried out. Both of these procedures yielded no close homologs. So, we opted to model the coat protein using homology domain modeling and remote-based homology modeling.

2.6. Disorderness Prediction. In order to characterize regions of sequences in coat protein which can be efficiently modeled, disorderness prediction was made. Disordered residue was identified using DISOPRED server (http://bioinf.cs.ucl.ac.uk/disopred/) [27] with a filter threshold of 5% and a false positive threshold of 2%. The disorderness is predicted by scanning the available sequence records in the PDB and then matches the electron density map to identify the missing coordinates. As a result, atomic coordinates of such amino acids will not be available for modeling of the protein and has the greater possibility of producing an irregular loop region in the modeled coat protein. Thus, manual search (only at the N-and C-terminals) for disordered sequence window in DISOPRED predictions was performed with the intention of excluding the corresponding region for modeling the coat protein. It was also ensured that disordered residue reported in the intervening sequence positions was left out so that the structure model did not possess any gaps.

2.7. Homology Domain Modeling. Robetta server (http://robetta.bakerlab.org/) was used for modeling the coat protein in which Ginzu, a hierarchical domain parsing and modeling protocol was adopted [28]. The input sequence (coat protein excluded with disorderness) was initially searched using BLAST, PSI-BLAST, FFA03 (http://ffas.burnham.org/), and 3D-Jury (http://meta.bioinfo.pl/) to obtain information on homologous regions which are then modeled with their comparative modeling protocol. Unassigned (i.e., nonhomologous as identified in the first stage) regions were then parsed to model as domain linkers using a combination of approaches, namely, HMMER search (http://hmmer.janelia.org/) over Pfam-A database and an MSA (produced from initial PSI-BLAST results) based search over NCBI NR database. Subsequently, K∗Sync alignment method was utilized to predict elements that are obligated to the fold to produce a single default alignment by dynamic programming. Best five models were generated after loop modeling, domain assembly, and side-chain packing.

2.8. Remote-Based Homology Modeling. Efforts were also carried out to predict structure of coat protein using remote-based homology modeling approach with the help of protein homology/analogy recognition engine (Phyre) version 2.0 (http://www.sbg.bio.ic.ac.uk/phyre2/) [29]. In the first step, profile was constructed using five iterations of PSI-BLAST against NR sequence database. The query secondary structure was predicted using three independent prediction programs (PSI-PRED (http://bioinf.cs.ucl.ac.uk/psipred/), SSPro (http://download.igb.uci.edu/sspro4.html) and Jnet (http://www.compbio.dundee.ac.uk/Software/JNet/jnet.html)) and a consensus prediction was made consequently. This profile and secondary structure were then scanned against the fold library using a profile-profile alignment algorithm to generate 3D models. Followed by a reconstruction procedure in the last stage, side-chains are packed using rotamer library and best models (selected based on confidence and sequence coverage) were returned.

2.9. Energy Minimization and Structure Validation of Models. The modeled structures were energy minimized using a utility in Tripos Benchware 3D Explorer (academic version; Tripos: A Certara company, http://www.tripos.com/) [30] with AMBER7 force field. Modeled structures were then validated for structure correctness and stereochemistry using Ramachandran plot [31] from RAMPAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) [32]. Based on the percentage of favournness and frequency of outliers, the models were selected and used for further analysis.

2.10. Prediction of Sequence- and Structure-Based DNA Binding Properties. In vitro studies showed that coat proteins from Geminivirusidae family bind nonspecifically with both ss- and ds-viral DNA [21, 33, 34]. To elucidate the role of DNA binding abilities of coat protein, sequence- and structure-based approaches were used. BindN (http://bioinfo.ggc.org/bindn/) employs support vector machines (SVMs) trained from data instances such as side chain pKₐ value, hydrophobicity index, and molecular mass of an amino acid [35]. A specificity of 80% with a filter threshold of 5% was chosen to avoid overwhelmed predictions. PreDs (http://pre-s.protein.osaka-u.ac.jp/preds/) makes use of molecular surface to evaluate electrostatic potential, local, and global curvatures of the PDB queried structure to predict potential dsDNA binding sites [36]. The modeled coat protein was defined as input with validation chosen from scoring functions.

2.11. Viral DNA Structure Modeling and Docking with Coat Protein. Canonical viral DNA was modeled using 3D-DART (3DNA-Driven DNA Analysis and Rebuilding Tool; http://haddock.chem.uu.nl/dna/dna.php) web service with default introduction of parameters for bends (roll, tilt, and twist) [37]. It uses 3DNA “fiber” module to generate canonical DNA structure and “find_pair” and “analyze” modules to produce a corresponding base pair (step) parameter file. The parameter file was used to set up local and global bends in the DNA structure file which are then remodeled finally using “rebuild” component to return PDB formatted DNA structure file. The docking phase was carried out using HADDOCK (High Ambiguity Driven biomolecular Docking; http://haddock.chem.uu.nl/) program [38] with the modeled coat protein and ds-viral DNA as inputs. Residues encompassed in a DNA binding region predicted by PreDs
with a Parea of greater than 250 Å² was specified as active site residues whereas passive residues were automatically defined around the active site which forms the boundary of the DNA-binding region. This specification was introduced to enhance the conformational search space for docking simulation as well as to avoid blindfold docking experiments. Definition of residues takes the form of experimental data which were converted into ambiguous interaction restraints (AIRs) in order to generate topology of the structures subsequently. The docking procedure consists of three stages: an energy minimization in a rigid-body manner, a semiflexible refinement in torsional space, and a final refinement in explicit solvent. After execution of each of these stages, the resultant structures are scored and ranked and the best fitted structures are employed in next stages. The best docked conformation can be obtained (usually clustered at the top) by inspecting the HADDOCK score which is a summation of intermolecular energies, namely, van der Waals (vdW), electrostatic (Elec), desolvation (Dsolv) and AIRs together with buried surface area (BSA): rigid-body score = 1.0 * Elec + 1.0 * vDW − 0.05 * BSA + 1.0 * Dsolv + 1.0 * AIR; final score = 1.0 * Elec + 1.0 * vDW + 1.0 * Dsolv + 1.0 * AIR.

2.12. Generation of Electrostatic Potential Map for Docked Structures. The influence of electrostatics for enabling DNA-protein interaction was studied using continuum Poisson-Boltzmann (PB) electrostatic approach. It was achieved by PBEQ-Solver (PBEQuation-Solver; http://www.charmmgui.org/?doc=input/pbeqsolver) [39] for which PQR files were required as molecular inputs. Hence, the docked complexes in PDB format were converted into PQR format using PDB2PQR server (http://www.pdb2pqr.org) [40]. PQR format embodies the replacement of occupancy column in a PDB file ("B") with the atomic charge ("Q") and the temperature factor column with the atomic radius ("R"). The inputted PDB file was subjected to following structural manipulations: rebuilding missing heavy atoms, building and optimizing hydrogens and assignment of atomic charges and radii based on force field parameters from CHARMM22 (selected option), AMBER99 or PARSE. All the PB calculations on PBEQ-Solver were performed in a coarse grid space (before focusing = 1.5 Å and after focusing = 1.0 Å) and utilized molecular surface (computed with a probe radius of 1.4 Å) to set up the dielectric boundary. The resultant electrostatic potential grid map in data explorer (dx) format was recovered and specified as input to PyMol version 2.5 program (academic version; Schrodinger LLC) [41] to view the PBEQ electrostatic map.

3. Results and Discussion

3.1. Prediction of Protein Family of Coat Protein. The protein family of ToLCV coat protein was predicted using a combination of programs. NCBI CD-Search using protein sequence revealed that it belongs to Gemini-coat protein superfamily (Pfam entry: pfam00844, accession no. Q8QYY9). Upon carefully examining the sequence alignment generated (E-value: 5.53e-100; bit score: 290.36) with SqLCV BR1 nuclear shuttle protein, it was studied that ToLCGV belongs to nuclear export factor BR1 family (Figure 1). BR1 is a ssDNA binding protein that shuttles between the nucleus and cytoplasm in plant cells [33].

PSI-BLAST sequence hit (PSI-BLAST threshold: 0.005 maximum iterations: 7; E-value: 1e-105; bit score: 383; sequence coverage in alignment: 99.64%) with a capsid protein of Begomovirus taxa (UniRef90 P03560; tomato golden mosaic virus) was observed. Further, sequence-based query over Pfam-A (Pfam-B not chosen as we focused on obtaining highly curated data) database produced a result similar to NCBI CD-Search. This HMM-based search provided an alignment with an E-value of 2.3e-87 and bit score of 292.1 (Figure 1). Manual inspection of PubMed references in the pfam00844 entry in NCBI CDD disclosed that coat proteins of Geminiviridae family binds ss- and ds-viral DNA in vitro [42]. For instance, TYLCV coat protein [43], maize streak virus (MSV) coat protein [21], SqLCV nuclear shuttle protein [44], and bean dwarf mosaic geminivirus (BDMV) movement protein [34] have the same function of binding which helps them to establish infection by nuclear shuttling of viral DNA across cell boundaries. Besides, coat protein also possesses binding function necessary for encapsidation of viral DNA. It is well known that the genomic component DNA-B in bipartite Begomovirus such as ToLCNDV encodes two movement proteins, namely, nuclear shuttle protein.

Figure 1: Sequence alignment of ToLCGV-[Vad] coat protein with nuclear export factor of BR1 family (Pfam entry: 00844 recovered from NCBI CDD) and HMM profile of the geminivirus coat protein.
3.2. Prediction of NLS in the Coat Protein Sequence. Mutagenesis study on MSV coat protein [4] and TYLCV [5] NLS region resulted in the cytoplasmic accumulation of the mutant protein. Thus, ToLCV coat protein must possess a NLS region in its sequence in order to be translocated to nucleus. A NLS signal was predicted with a score 10.2 by the cNLS Mapper in the coat protein N-terminal with a composition of 20 amino acids (predicted bipartite NLS: MSKRPADMLIFTASKYRRR, predicted monopartite NLS: none).

The occurrence of basic amino acids in the predicted NLS showed that lysine and arginine constituted 2 and 4 counts which proposed to have a classic NLS pattern (Table 1) and can be comparable to experimentally identified TYLCV coat protein NLS [5] (Figure 2(a)). Despite the impressive number of receptor-cargo interactions that have been studied, the prediction of NLSs in candidate proteins remains extremely difficult. So, we step forwarded our search in scientific literatures related to BR1 nuclear export family in order to infer the predictions made. Pairwise sequence alignment of NLS region from MSV and ToLCGV-[V] coat proteins resulted in an identity and similarity percentage of 25% and 33.3% with a score of −6 whereas TYLCV and ToLCGV-[Vad] yielded an identity and similarity percentage of 51.7% and 58.6% with a score of 49. This pairwise alignment suggested that ToLCV coat protein is much more conserved with TYLCV rather than SqLCV [3] (identity and similarity = 15%, score = −20), a representative protein member of BR1 family (pfam00844) in Pfam database.

3.3. MSA and Phylogenetic Analysis of Indian Strains. ToLCV coat protein sequences from Indian strains were retrieved from NCBI database to construct MSA in order to study amino acids crucial in conserved domain and responsible for systemic infection, viral particle formation, and insect transmission. Norris et al. 1998 reported that a functional coat protein having amino acids in the following sequence positions, namely, Pro/Gln129, Gln/His134, and Glu/Asp152 on TYLCV isolates is essential for correct assembly of virions and transmission by the insect vector [47]. These key residues were identified by B. tabaci transmissibility studies in the field isolates of TYLCV-Sic (Sicily), TYLCV-Sar (Sardinia), and TYLCV-SicRv (engineered mutant of Sicily) [47]. Examination of corresponding positions in our MSA cluster revealed that Lys129, Ser/Thr134, and Asp151 (instead of 152nd position as a result of single residue deletion) were
and the disorderness as the corresponding sequence position
minal. There exists a relationship between the predicted NLS
was scattered with peaks demonstrating residue disorderness
showed that a window with sequence positions from 1 to 50
terminals. Disorder profile produced with 5% filter threshold
was taken to eliminate the loop region in the sequence
protein modeling procedures adopted by us. This scrutiny
sequence regions that cannot be modeled e
orderness was predicted in the coat protein to identify the
3.4. Disorderness and Their Link with Predicted NLS. Dis-
conserved amongst Indian strains in comparison to non-
Indian isolates and are found to be wild-type. The com-
parison of chemical properties of the template with that of
MSA showed that a positively charged amino acid (Lys129)
was identified in the uncharged polar (Gln129) position.
The second important residue (Ser/Thr134 in replacement
with Gln/His134) was conserved in terms of polarity while a
negatively charged residue (Glu/Asp152; Asp151) was pre-
served in the third crucial position (Figure 2(b)). This amino
acids combination (Lys129, Ser/Thr134, and Asp121) is also
conserved in coat proteins among different wild-type viruses,
namely, tomato golden mosaic virus, tomato mottle virus-
[Florida], sinaloatomato leaf curl virus, tomato leaf crumple
virus, taino tomato mottle virus, abutilon mosaic virus-
[Hawaii], bean golden mosaic virus-[Brazil], SqLCV and
papaya leaf curl virus [47].

A phylogenetic tree based on NJ algorithm was con-
structed using Phylip version 3.68 to study the sequence
conservativeness among Indian strains. Surprisingly, coat
proteins characteristic from districts, namely, Vadodara,
Varanasi, and Kelloo were clustered in a node with a branch
length of 0.167. It should be noted that these members were
representing different states in the Northern India contrast-
ing to other members which were sufficiently diverged to
each other. Besides the fact that Bangalore isolates were con-
served among each other, they were distinct with one of the
state member, Karnataka with a length of 0.011. Isolates from
New Delhi and Lucknow were conserved as expected in terms
of area nearness (Figure 3). The key amino acids required
for biochemical functions were indeed conserved amongst
each other with respect to the comparison using MSA made
above.

### Table 1: Known and predicted NLS pattern of BR1 nuclear export family.

| Organism            | Predicted NLS pattern                        | Frequency of lysine residues | Frequency of arginine residues |
|---------------------|---------------------------------------------|-----------------------------|-------------------------------|
| SqLCV coat protein  | KRSYGAARGDDRRRP (Sanderfoot et al., 1996 [3]) | 1                           | 5                             |
| MSV coat protein    | MSTSKKRRDDSNWSKRVTKKKPS (Liu et al., 1999 [4]) | 6                           | 3                             |
| TYLCV coat protein  | MSKRPDIIISTPVSVKRRLNFSDPYSS (Kunik et al., 1998 [5]) | 2                           | 4                             |
| ToLCGV-[Vad]       | MSKRPAFLIPASKVRRR (predicted by cNLS Mapper) | 2                           | 4                             |

3.5. Structure Modeling of Coat Protein. No close template
was predicted as loop region with a variety of secondary
structure prediction programs including PSI-Pred, GORIV,
and so forth. So, we decided to exclude NLS signal from the
protein sequence for modeling due to the consideration of
disordered profile and the increased possibility of generating
loop geometry.

3.6. Selection of Best Scoring Models. A total of six coat pro-
tein models (five obtained from Robetta program and one
from Phyre program) were subjected to energy minimization
with AMBER7 force field and 250 as maximum number of
evaluations using minimize energy module engineered in
Tripos Benchware 3D Explorer. Energy minimized structures
were then validated using stereochemistry check with the
help of Ramachandran plot. The Robetta model (energy
minimized to −2682.00 Kcal/mol) was chosen based on
the residue occurrence of more than 95% estimated by
summing up favorable and allowed regions (Figure 5) in ϕ-ψ
core areas and the presence of only one outlier (Glu204) whereas the best scored Phyre model was discarded due to the loose packing of loop regions and its close resemblance to Robetta model (root mean squared deviation (RMSD): 20.8578 Å over 1041 matched atoms). There is one more reason to unconsider the Phyre model as the N-terminus was constituted with loop elements, two helices in the intervening region (two helices expected at the N-terminal) and eleven β-strands (only eight were expected instead of eleven) as it was not complied with characteristic Geminate viral particle. So, we discarded these models for further analysis.

3.7. Resemblance to Geminate Viral Particle. The modeled ToLCV coat protein was compared with the structure of the MSV Geminate particle, a member of Geminiviridae family determined using cryo-electron microscopy and three-dimensional image reconstruction methods [49]. The modeled protein possessed an N-terminal helix with an eight-stranded antiparallel β-barrel motif characteristic of Geminate particle. The β-barrel motif is a dominant structural unit in all ssDNA virus structures that have been determined to atomic resolution. Unlike the model generated by Zhang et al. 2001 [49], our Robetta model has 7 reliable and 1 short (total 8 strands) antiparallel β-strands and 2 helices at the N-terminal instead of 1 in comparison to Zhang’s Geminate model. We also noticed 1 helix at the intervening region (Figure 6). We expected that this additional accumulation of secondary structural elements is beyond the evolution and may be an additional procurement in the dsDNA virus structures in disparity to ssDNA virus structures due to conservational pressures as described by Zhang et al. 2001 [49] or due to the insertion sequences or it may be due to the loop geometry as it was undistinguished by the present

Figure 3: Phylogenetic tree of ToLCV coat protein representing closeness among Indian strains.

Figure 4: Disordered profile of ToLCV coat protein with disorderness in the sequence positions 1–50 which contains predicted NLS.
3.8. Sequence and Structure-Based DNA Binding Properties.
ToLCV must possess DNA binding properties for accomplishing several cellular functions such as nuclear targeting of viral DNA, encapsidation of viral DNA, systemic infection, viral particle formation, insect transmission, and correct assembly of virions as experimentally studied in other members of the family, namely, MSV, SqLCV, BDMV, and so forth. Interestingly, the coat proteins bind both ss- and ds-viral DNA in sequence independent fashion [21, 33, 34]. To reveal the crucial amino acids involved in DNA interaction, sequence- and structure-based approaches were utilized. BindN predicted spatially distributed residues as component of DNA binding interface with 74 amino acids and 28.90% contribution (Figure 7). We expected that this widespread distribution of DNA binding residues will come together during protein folding and will interact with viral DNA. Thus, we step forwarded to identify those prominent amino acids using our generated ToLCV coat model. PreDs revealed loop regions as potential DNA binding region accompanied with all those amino acids predicted by BindN. This prediction was validated by inspecting the scoring functions such as
Pscore and Parea. Pscore is an indicator for the ratio of the predicted area possessing maximum value while Parea represents area of the predicted ds-DNA binding region on the protein surface. We achieved 0.31 as Pscore (threshold: >0.12) and 2102.26 Å² (threshold: >250 Å²) as Parea, respectively (Figure 8).

The TYLCV coat protein gets imported into the plant and insect cells nuclei via using its N-terminal NLS [5]. It has also been proposed that TYLCV coat protein functions as BR1 protein facilitating DNA trafficking across cell boundaries and demonstrated that the coat protein binds DNA cooperatively [43]. It is also highlighted that TYLCV coat protein may also aid in protecting the transported coat protein-DNA complex from intracellular nucleolytic degradation as this complex was highly resistant to in vitro S1 nuclease activity [43]. In MSV coat protein, the DNA binding domain was mapped to the N-terminal 104 amino acids inclusive of NLS [21]. Immuno-electron microscopy revealed that DNA binding domain between residues 5 and 22 suggested that this region could be involved in transporting geminivirus coat protein towards nuclei [50]. We predicted that certain N-terminal amino acids of ToLCV coat protein scored a confidence of 6–9 indicating DNA binding properties (Figure 7). Besides, structural analysis of modeled ToLCV indicated that N-terminal residues contributing to loop secondary structure form the major element in interacting with viral DNA as described below.

### 3.9. Viral DNA Structure Modeling and Docking with Coat Protein

We developed a canonical ds-viral DNA using 3D-DART with generalized geometrical constraints to explore the interaction with coat protein. Molecular docking was performed using HADDOCK program with predicted DNA binding region as active site. Best scoring clusters were sorted based on HADDOCK score. The most reliable (top) cluster having four similar docked conformations (HADDOCK score: −12.0 ± 11.0) were recovered. The DNA binding interface of coat protein was found to be the loop region whereas major groove was the molecular interface of viral DNA in which the best conformers were sampled (Figure 9). The intermolecular energies (unit in KJ/mol) obtained are as follows: vDW: −66.9 ± 5.5, Elec: 827.4 ± 87.8, Dsolv: 125.7 ± 11.7, AIR: 947.0 ± 50.09, and BSA: 2099.6 ± 137.9. We also observed that RMSD of overall lowest-energy structure was 4.3 ± 3.0 with respect to structures in different clusters. The internal energy in apo form (free molecules) was 278722.00 KJ/mol whereas in bound form (DNA-protein complex) was 15494.00 KJ/mol and the binding energy was predicted as −264139.00 KJ/mol.
3.10. Electrostatic Interaction of Coat Protein. DNA binding interface predicted by PreDs was found to be electrostatically favored as its prediction was principally based on electrostatic potential, local, and global curvatures present on the DNA surface. HADDOCK revealed coat protein’s loop topology as its molecular interface unit. DNA-protein interaction is predominantly influenced by electrostatics and can be efficiently studied using adaptive Poisson-Boltzmann solver (APBS) approach. The docked conformation was charged appropriately based on physiological environment with preprocessing using PDB2PQR and all PB calculations were carried out using PBEQ-Solver in a coarse grid space. Electrostatic grid map analyzed using PyMol showed that neutral patches in coat protein were found to interact with viral DNA (Figure 10). We expected that an isosurface of positive patch will tend to interact with negatively charged viral DNA. Upon manual inspection of charged clusters in the coat protein, the positive charged clusters were cornered in surface with significantly low negative patches and a greater deal of positive regions corresponded to β-barrel motif. We ruled out the requirement of positive patch contributing to DNA interaction as geometrical flexibility of loop region in the coat protein having neutrality was much favored rather than charge-charge attraction. This view was further inferred by the best docked conformation in which electrostatic energy term was investigated and found to be the major intermolecular descriptor representing interaction.

Zrachya et al. 2007 reported that TYLCV coat protein could be targeted by small interfering ds RNAs (siRNAs) derived from intron-hairpin RNA construct to develop disease resistant transgenic tomato cultivars and showed its potential in N. benthamiana transient assays [51]. Similar studies targeting against antisense replicase gene (AC1) in ToLCV helped in developing trait-stable transgenics [52]. Besides, the midsized aggregation of coat protein inside nucleus is associated with resistance whereas large aggregates leading to infection susceptibility [53]. We highlighted sequence regions in ToLCV coat protein possessing both DNA binding properties and the functional amino acids combination essential for virulence and these regions of interest can be targeted by developing siRNA. In addition, these molecular properties of ToLCV coat protein can be accounted in developing Begomovirus resistant-engineered plants.

4. Conclusion

ToLCV coat protein possesses DNA binding properties to function similar to BR1 nuclear shuttle. The amino acid combinations crucial for virulence were investigated through MSA and the evolutionary relationship was traced by phylogenetic analysis which indicated that the Indian strains are closely related in the context of geographical locations. The predicted NLS of ToLCV coat protein shares more similarity with experimentally known TYLCV coat protein NLS. Molecular modeling represented ToLCV coat protein as Geminate viral particle. Further, sequence- and structure-based approaches identified that ToLCV coat protein through its loop topology interacts with viral DNA as surface complementarity proven to be the major promoting factor followed by electrostatic interaction. We anticipate the conserved region of ToLCV coat protein prominent for DNA binding and functional sequence pattern can be targeted by RNA interference to develop disease-resistant transgenic tomato plants.

Acknowledgments

S. P. Kumar acknowledges generous support from DST, New Delhi as Innovation in Science Pursuit for Inspired Research (INSPIRE) Fellowship. The authors would like to thank three anonymous reviewers for their fruitful suggestions.
References

[1] B. Gronenbor, “The Tomato yellow leaf curl virus genome,” in Tomato Leaf Curl Virus Disease: Management, Molecular Biology, Breeding For Resistance, H. Czosnek, Ed., vol. 8, pp. 67–84, Springer, The Netherlands, 2007.

[2] M. J. Melzer, D. Y. Ogata, S. K. Fukuda et al., “First report of Tomato yellow leaf curl virus in Hawaii,” Plant Disease, vol. 94, no. 5, pp. 641–643, 2010.

[3] A. A. Sanderfoot, D. J. Ingham, and S. G. Lazarowitz, “A viral movement protein as a nuclear shuttle: the geminivirus BR1 movement protein contains domains essential for interaction with BL1 and nuclear localization,” Plant Physiology, vol. 110, no. 1, pp. 23–33, 1996.

[4] H. Liu, M. I. Boulton, C. L. Thomas, D. A. M. Prior, K. J. Oparka, and J. W. Davies, “Maize streak virus coat protein is karyophyllic and facilitates nuclear transport of viral DNA,” Molecular Plant-Microbe Interactions, vol. 12, no. 10, pp. 894–900, 1999.

[5] T. Kunik, K. Palanichelvam, H. Czosnek, V. Citovsky, and Y. Gafni, “Nuclear import of the capsid protein of Tomato yellow leaf curl virus (TYLCV) in plant and insect cells,” Plant Journal, vol. 13, no. 3, pp. 393–399, 1998.

[6] S. Chakraborty, P. K. Pandey, M. K. Banerjee, G. Kalloo, and C. M. Fauquet, “Tomato leaf curl Gujarat virus, a New begomovirus species causing a severe leaf curl disease of tomato in Varanasi, India,” Phytopathology, vol. 93, no. 12, pp. 1485–1495, 2003.

[7] R. W. Briddon, J. K. Brown, E. Moriones et al., “Recommendations for the classification and nomenclature of the DNA-β satellites of Begomoviruses,” Archives of Virology, vol. 153, no. 4, pp. 763–781, 2008.

[8] A. Kheyr-Pour, M. Bendahmane, V. Matzeit, G. P. Accotto, S. Crespi, and B. Gronenborn, “Tomato yellow leaf curl virus from Sardinia is a whitefly-transmitted monopartite geminivirus,” Nucleic Acids Research, vol. 19, no. 24, pp. 6763–6769, 1991.

[9] P. Jyothsna, R. Rawat, V. G. Malathi, “Predominance of Tomato leaf curl Gujarat virus as a monopartite Begomovirus: association with Tomato yellow leaf curl Thailand betasatellite,” Archives of Virology. In press.

[10] P. Pandey, S. Mukhopadhy, A. R. Naqui, S. K. Mukherjee, G. S. Shekhawat, and N. R. Choudhury, “Molecular characterization of two distinct monopartite Begomoviruses infecting tomato in India,” Virology Journal, vol. 7, article no. 337, 2010.

[11] T. Kunik, R. Salomon, D. Zamir et al., “Transgenic tomato plants expressing the Tomato yellow leaf curl virus capsid protein are resistant to the virus,” Nature Biotechnology, vol. 12, no. 5, pp. 500–504, 1994.

[12] S. K. Raj, R. Singh, S. K. Pandey, and B. P. Singh, “Agrobacterium-mediated tomato transformation and regeneration of transgenic lines expressing Tomato leaf curl virus coat protein gene for resistance against TLCV infection,” Current Science, vol. 88, no. 10, pp. 1674–1679, 2005.

[13] D. Pratap, S. K. Raj, S. Kumar, S. K. Snehi, K. K. Gautam, and A. K. Sharma, “Coat protein-mediated transgenic resistance in tomato against a IB subgroup Cucumber mosaic virus strain,” Phytoparasitica, vol. 40, no. 4, pp. 375–382, 2012.

[14] S. Chakraborty, R. Vanitharan, B. Chattopadhyay, and C. M. Fauquet, “Supervirulent pseudorecombination and asymmetric synergism between genomic components of two distinct species of Begomovirus associated with severe tomato leaf curl disease in India,” Journal of General Virology, vol. 89, no. 3, pp. 818–828, 2008.

[15] N. Kirthi, S. P. Maiya, M. R. N. Murthy, and H. S. Savithri, “Evidence for recombination among the Tomato leaf curl virus strains/species from Bangalore, India,” Archives of Virology, vol. 147, no. 2, pp. 255–272, 2002.

[16] D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler, “GenBank,” Nucleic Acids Research, vol. 34, pp. D16–D20, 2006.

[17] A. Marchler-Bauer and S. H. Bryant, “CD-Search: protein domain annotations on the fly,” Nucleic Acids Research, vol. 32, pp. W327–W331, 2004.

[18] S. F. Altschul, T. L. Madden, A. A. Schäffer et al., “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs,” Nucleic Acids Research, vol. 25, no. 17, pp. 3389–3402, 1997.

[19] A. Bateman, E. Birney, L. Cerruti et al., “The Pfam protein families database,” Nucleic Acids Research, vol. 30, no. 1, pp. 276–280, 2002.

[20] S. Kosugi, M. Hasebe, T. Entani, S. Takayama, M. Tomita, and H. Yanagawa, “Design of peptide inhibitors for the importin α/β nuclear import pathway by activity-based profiling,” Chemistry and Biology, vol. 15, no. 9, pp. 940–949, 2008.

[21] H. Liu, M. I. Boulton, and J. W. Davies, “Maize streak virus coat protein binds single- and double-stranded DNA in vitro,” Journal of General Virology, vol. 78, no. 6, pp. 1265–1270, 1997.

[22] P. Rice, L. Longden, and A. Bleasby, “EMBOSS: the European molecular biology open software suite,” Trends in Genetics, vol. 16, no. 6, pp. 276–277, 2000.

[23] M. A. Larkin, G. Blackshields, N. P. Brown et al., “Clustal W and Clustal X version 2.0,” Bioinformatics, vol. 23, no. 21, pp. 2947–2948, 2007.

[24] J. Felsenstein, “PHYLIP- phylogeny inference package (version 3.2),” Cladistics, vol. 5, no. 2, pp. 164–166, 1989.

[25] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, and T. L. Madden, “NCBI BLAST: a better web interface,” Nucleic acids research, vol. 36, pp. W5–9W, 2008.

[26] F. C. Bernstein, T. F. Koetzle, and G. J. B. Williams, “The protein data bank: a computer based archival file for macromolecular structures,” Journal of Molecular Biology, vol. 112, no. 3, pp. 535–542, 1977.

[27] J. J. Ward, L. J. McGuffin, K. Bryson, B. F. Buxton, and D. T. Jones, “The DISOPRED server for the prediction of protein disorder,” Bioinformatics, vol. 20, no. 13, pp. 2138–2139, 2004.

[28] D. E. Kim, D. Chivian, and D. Baker, “Protein structure prediction using the Robetta server,” Nucleic Acids Research, vol. 32, pp. W526–W531, 2004.

[29] L. A. Kelley and M. J. Sternberg, “Protein structure prediction on the web: a case study using the Phyre server,” Nature protocols, vol. 4, no. 3, pp. 363–371, 2009.

[30] Tripos, A Certara Company, “Tripos benchware 3D explorer for interactive molecular visualizations and structure manipulations,” St. Louis, Miss, USA, http://www.tripos.com/.

[31] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, “Stereochemistry of polypeptide chain configurations,” Journal of Molecular Biology, vol. 7, pp. 95–99, 1963.

[32] S. C. Lovell, I. W. Davis, W. B. Arendall et al., “Structure validation by Cα geometry: ϕ, ψ and Cβ deviation,” Proteins, vol. 50, no. 3, pp. 437–450, 2003.

[33] E. Pascal, A. A. Sanderfoot, B. M. Ward, R. Medville, R. Turgeon, and S. G. Lazarowitz, “The geminivirus BR1 movement protein binds single-stranded DNA and localizes the cell nucleus,” Plant Cell, vol. 6, no. 7, pp. 995–1006, 1994.

[34] M. R. Rojas, A. O. Noueiry, W. J. Lucas, and R. L. Gilbertson, “Bean dwarf mosaic Geminivirus movement proteins recognize...
DNA in a form- and size-specific manner,” *Cell*, vol. 95, no. 1, pp. 105–113, 1998.

[35] L. Wang and S. J. Brown, “BindDN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences,” *Nucleic Acids Research*, vol. 34, pp. W243–W248, 2006.

[36] Y. Tsuchiya, K. Kinoshita, and H. Nakamura, “PreDs: a server for predicting dsDNA-binding site on protein molecular surfaces,” *Bioinformatics*, vol. 21, no. 8, pp. 1721–1723, 2005.

[37] M. van Dijk and A. M. Bonvin, “3D-DART: a DNA structure modelling server,” *Nucleic Acids Research*, vol. 37, no. 2, pp. W235–W239, 2009.

[38] S. J. de Vries, M. van Dijk, and A. M. Bonvin, “The HADDOCK web server for data-driven biomolecular docking,” *Nature protocols*, vol. 5, no. 5, pp. 883–897, 2010.

[39] S. Jo, M. Vargyas, J. Vasko-Szedlar, B. Roux, and W. Im, “PBEQ-Solver for online visualization of electrostatic potential of biomolecules,” *Nucleic acids research*, vol. 36, pp. W270–W275, 2008.

[40] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker, “PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations,” *Nucleic Acids Research*, vol. 32, pp. W665–W667, 2004.

[41] M. A. Lill and M. L. Danielson, “Computer-aided drug design platform using PyMOL,” *Journal of Computer-Aided Molecular Design*, vol. 25, no. 1, pp. 13–19, 2011.

[42] A. O. Noueiry, W. J. Lucas, and R. L. Gilbertson, “Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport,” *Cell*, vol. 76, no. 5, pp. 925–932, 1994.

[43] K. Pananichvilam, T. Kunik, V. Citovsky, and Y. Gafni, “The capsid protein of *Tomato yellow leaf curl virus* binds cooperatively to single-stranded DNA,” *Journal of General Virology*, vol. 79, no. 11, pp. 2829–2833, 1998.

[44] S. Qin, B. M. Ward, and S. G. Lazarowitz, “The bipartite geminivirus coat protein aids BR1 function in viral movement by affecting the accumulation of viral single-stranded DNA,” *Journal of Virology*, vol. 72, no. 11, pp. 9247–9256, 1998.

[45] M. Hussain, S. Mansoor, S. Iram, A. N. Fatima, and Y. Zafar, “The nuclear shuttle protein of *Tomato leaf curl New Delhi virus* is a pathogenicity determinant,” *Journal of Virology*, vol. 79, no. 7, pp. 4434–4439, 2005.

[46] M. R. Rojas, H. Jiang, R. Salati et al., “Functional analysis of proteins involved in movement of the monopartite Begomovirus, *Tomato yellow leaf curl virus*,” *Virology*, vol. 291, no. 1, pp. 110–125, 2001.

[47] E. Noris, A. M. Vaira, P. Caciagli, V. Masenga, B. Gronenborn, and G. P. Accotto, “Amino acids in the capsid protein of *Tomato yellow leaf curl virus* that are crucial for systemic infection, particle formation, and insect transmission,” *Journal of Virology*, vol. 72, no. 12, pp. 10050–10057, 1998.

[48] T. A. Jones and L. Liljas, “Structure of satellite tobacco necrosis virus after crystallographic refinement at 2.5 Å resolution,” *Journal of Molecular Biology*, vol. 177, no. 4, pp. 735–767, 1984.

[49] W. Zhang, N. H. Olson, T. S. Baker et al., “Structure of the maize streak virus geminate particle,” *Virology*, vol. 279, no. 2, pp. 471–477, 2001.

[50] M. S. Pinner, V. Medina, K. A. Plaskitt, and P. G. Markham, “Viral inclusions in monocotyledons infected by Maize streak and related viruses,” *Plant Pathology*, vol. 42, no. 1, pp. 75–87, 1993.

[51] A. Zrachya, P. P. Kumar, U. Ramakrishnan et al., “Production of siRNA targeted against TYLCV coat protein transcripts leads to silencing of its expression and resistance to the virus,” *Transgenic Research*, vol. 16, no. 3, pp. 385–398, 2007.
