In silico Analysis of Genetic Diversity of Begomovirus using Homology Modelling

Department of Biotechnology, MITS University, Lakshmangarh, Rajasthan, India

Abstract: Structural bioinformatics is concerned with computational approaches to predict and analyse the spatial structure of proteins and nucleic acids. The 3D models can be obtained most easily if the 3D structure of a homologous protein is known (homology modeling, comparative modeling). Homology modeling or comparative modeling is based on the fact that a structure of a protein can be reliably modeled when its sequence is sufficiently similar to a protein sequence with known 3D structure. The Geminivirus taxonomic group of plant viruses is characterized by geminate particles and genomes consisting of single-stranded circular DNA (sscDNA) molecules of about 2.5 to 2.8 kb in size. Here, we analyzed the genetic diversity of begomovirus from different crops and weeds of different region on the basis of homology modeling. This high degree of genetic diversity suggests a recent transfer of indigenous begomovirus from wild hosts into economic important crops.

Key words: Homology modeling, geminivirus, genetic diversity, geminate particles

INTRODUCTION

Today, experimentally we can only predict the structure-based drug design, analysis of protein function, interaction, antigenic behaviour and rational design of protein with increased stability or novel function. Unlike experimental structure, Protein modeling are used to predict a structure from its sequence with an accuracy that allow users to use safely generated in silico protein models in all the contexts. In addition, protein modeling is the only way to obtain structural information if experimental techniques fail. Many proteins are too large for NMR analysis and cannot be crystallized for X-ray diffraction.

The UCLA-DOE Structure Evaluation server is a tool designed to help in the refinement of crystallographic structures. It will provide you with a visual analysis of the quality of a putative crystal structure for a protein. Verify 3D expects this crystal structure to be submitted in PDB format. Please note that Verify 3D works best on proteins with at least 100 residues. To submit a crystal structure for analysis, simply select it with the file dialog which is activated by clicking on the Browse button below, then click the Send File button.

Verify3D analyzes the compatibility of an atomic model (3D) with its own amino acid sequence (1D). Each residue is assigned a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar, etc). A collection of good structures is used as a reference to obtain a score for each of the 20 amino acids in this structural class. The scores of a sliding 21-residue window (from -10 to +10) are added and plotted for individual residues.

Genomic organization of begomovirus: The genus Begomovirus contains more than 200 species (Fauquet et al., 2008) and belongs to the taxonomic family Geminiviridae. They are plant viruses that as a group have a very wide host range, infecting dicotyledonous plants. Worldwide they are responsible for a large amount of economic damage to many important crops such as tomatoes, beans, squash, cassava and cotton.

Virus particles are non-enveloped. The nucleocapsid is 38 nm long and 15-22 nm in diameter. While particles have basic isosahedral symmetry, they consist of two incomplete isosahedra missing one vertex-joined together. There are 22 capsomeres per nucleocapsid.

The geminivirus taxonomic group of plant viruses is characterized by geminate particles and genomes consisting of single-stranded circular DNA (sscDNA) molecules of about 2.5 to 2.8 kb in size (Lazarowitz, 1987). There are currently three subgroups recognized by the International Committee on Taxonomy of Viruses (ICTV): subgroup I viruses have single component, ssDNA genomes of about 2-7 kb, mainly infect monocotyledonous plants and are obligate leafhopper-transmitted; Subgroup II single genomic component closely resembles DNA A of III, transmitted by leafhoppers or tree hoppers and infect dicotyledonous plant subgroup III viruses have two component genomes, infect dicotyledonous plants and are transmitted.

Corresponding Author: V.K. Gupta, Department of Biotechnology, MITS University, Lakshmangarh, Rajasthan, India
Tel: 00919352215170
Fig. 1: Genomic organization of a typical begomovirus

mechanically and by the whitefly *Bemisia tabaci* (ICTV Group Descriptions, 1989; G.P. Martelli, personal communication). The two subgroups are distantly evolutionarily related (Howarth and Vandemark, 1989).

Many begomoviruses have a bipartite genome, this means that the genome is segmented into two segments (referred to as DNA A and DNA B) that are packaged into separate particles. Both segments are generally required for successful symptomatic infection in a host cell, but DNA B is dependent for its replication upon DNA A, which can in some begomoviruses apparently cause normal infections on its own. Depending on the particular begomovirus, six different proteins are involved in mediating virus transport through host plants: CP, AV2, AC1, AC2, AC3 and AC4 (Fig. 1) (Rojas et al., 2005).

Unlike monopartite begomoviruses, the CP of bipartite begomoviruses is not required for cell-to-cell or long-distance movement. For bipartite begomoviruses, cell-to-cell movement function is carried out by two proteins encoded by the DNA B component, NSP and MP. In the case of monopartite begomoviruses, two proteins, V1 and C4, have been implicated in cell-to-cell movement function (Jupin et al., 1994; Rojas et al., 2001). The opposing transcription units of begomovirus DNA-A and -B molecules are separated by an Intergenic Region (IR) that generally shares a highly conserved region of approximately 200 nt, named the Common Region (CR) (Lazarowitz, 1992). The CR contains an origin of replication (ori) that includes a stem-loop structure containing an invariant nonnucleotide (TAATATAC) sequence whose T7-A8 site is required for cleaving and joining viral DNA during replication and conserved iterated sequences (iterons) required for specific recognition and binding by Rep during replication (Arguello-Astorga et al., 1994; Fontes et al., 1994a, b).

Based on phylogenetic studies and genome arrangement, begomoviruses have been divided broadly into two groups: the Old World (OW) viruses (eastern hemisphere, Europe, Africa, Asia) and the New World (NW) viruses (western hemisphere, the Americas) (Padidam et al., 1999; Paximadis et al., 1999, Rybicki, 1994). Begomovirus genomes have a number of characteristics that distinguish Old World and New World viruses. All indigenous New World begomoviruses are bipartite, whereas both bipartite and monopartite begomoviruses are present in the Old World. In addition, DNA-A of bipartite begomoviruses from the New World lacks an AV2 ORF (Rybicki, 1994; Stanley and Latham, 1992).

Besides this few geminiviruses contains a small circular single-stranded DNA termed satellite DNA or DNA-B having single open reading frame (ORF). DNA-B are biologically functional, being required for amplification of DNA A to high levels and for the efficient systemic spread of infection. Since they were first reported in 2000, over one hundred full-length DNA-Bs have been cloned and sequenced (Mansoor et al., 2006). They consist of approximately 1350 nt and depend on their helper begomoviruses for replication and encapsidation. The process through which DNA-B regulates pathogenesis is unknown, but it might affect the replication of its helper virus by either facilitating its spread in host plants, or by suppressing host gene silencing (Saunders et al., 2000). The DNA b-encoded protein, bc1, is the determinant of both pathogenicity and suppression of gene silencing (Bridgon and Stanley, 2006).

**Genetic diversity of ORF region**: AV1/AV2. DNA A of the begomovirus contains AVI ad AV2 at the virus strand. The localization of the product of this ORF in secondary plasmodesmata with the onset of viral lesions is consistent with its role in the movement of monopartite geminivirus (Dickinson et al., 1996). All begomoviruses form the new world are bipartite and lack a virion sense Open Reading Frame (ORF) designated AV2. This involved in the movement of monopartite viruses but its function is unknown in bipartite. A recent report has similarly shown that AV2 of Indian Cassava Mosaic Virus (ICMV), another bipartite begomovirus from the Indian subcontinent, is involved in virus spread to neighbouring cells (Shunter and Bisaro, 1992), whereas the V2 (a homolog of AV2) of a monopartite begomovirus is involved in overcoming host defences mediated by post-transcriptional gene silencing as well as in movement (Shunter et al., 2001; Vance and Vacheret, 2001). The comparison of 153 residues of two begomovirus namely Tomato yellow leaf curl Thailand virus and Tomato leaf curl Karnataka virus shows position diversity of aromatic amino acids as in Fig. 2. The Chl-Chl2 plot also shows the diversity of favourable and allowed region of the structure (Fig. 3).
**AC1:** The Rep is present in nuclei of infected plant cells (Nagar et al., 1995), where it plays a key role in geminivirus DNA replication and transcription (Laufs et al., 1995). Rep confers origin of replication and initiates plus-strands DNA replication (Chatterji et al., 2001). Several biochemical activities have been demonstrated for the Rep, specifically cleaves and ligates viral plus strands DNA within the loop of the hairpin motifs. The protein also possesses DNA nicking activity, site specific topoisomerase and helicase activity (Heyraud-Nitschke et al., 1995). Amino acids 1 to 130 are sufficient for DNA binding, amino acids 134 to 181 are for oligomerization activity. It is supposed that C-terminus is involved in ATP hydrolysis (Desbiez et al., 1995, Fig. 4).

**AC2/TrAP:** The AC2 gene product or transcriptional activator protein (TrAP) is a trans-acting factor needed for the expression of both AV1 (coat protein) and BV1 genes (Sunter and Bisaro, 1992). The TrAP required for initiating transcription of the virion-sense genes and also involved in suppression of post-transcriptional gene silencing-mediated host defences (Voinnet et al., 1999, Vanitharami et al., 2005). Southern blot analyses of the total DNAs isolated from inoculated protoplasts revealed that EACMCV AC2 expressed in trans under a 35S promoter enhanced ACMV-[CM] DNA accumulation by ~8-fold compared to that of protoplasts transfected with ACMV-[CM] (Vanitharami et al., 2004). ACMV-[CM] AC2 in trans did not alter EACMCV DNA accumulation, but it could act as a mild suppressor of PTGS, as shown for AC2 of ACMV-[KE] (Hamilton et al., 2002; Voinnet et al., 1999). The AC2 genes of ACMV-[CM] and EACMCV show 56% homology and contain conserved zinc finger, transactivator and DNA-binding domains; however, these genes act differently. In monopartite geminiviruses, it has been demonstrated recently that the C2 protein of TYLCCNV requires the DNA-binding domain, the zinc finger domain and the nuclear localization signal for anti-PTGS activity (Van Wezel et al., 2002, 2003). C2 is a positional homologue of AC2 in bipartite geminiviruses, but it varies functionally in that it does not facilitate the transcriptional activation of coat protein, as in the case of AC2 of bipartite geminiviruses. Therefore, we propose that the different functions of the AC2 proteins that we noticed could be due to differences at the level of point mutations, as indicated by sequence comparisons (data not shown). In addition, AC2 of Tomato golden mosaic virus and C2 of Beet curly top virus have been shown to manipulate host metabolism by interfering with SNF1 kinase and adenosine kinase to provide enhanced susceptibility to virus infection (Wang et al., 2003, Silhavy et al., 2002); however, their roles in relation to PTGS have not been established. All these data support the fact that these tiny proteins are multifunctional in nature.

**AC3/Ren:** The AC3 is also known as replication enhancer (ren) having the length of 134 nt. Base composition of AC3 shows that it has lower average GC content in comparison to other GC content found among other. It has been observed that ren protein is located in nuclei of infected plant cells at levels similar to Rep (Nager et al., 1995) suggesting that it might act with Rep during initiation of viral DNA replication. Experimental observation suggests that ren protein might increase the affinity of rep for the origin (Mohr et al., 1990). Rep-Ren protein complex formation might also enhance DNA cleavage, ligation or putative helicase (Koonin et al., 1991) activities of Rep.
Fig. 3: Chi1-Chi2 plot for AV1 coat protein of a begomovirus. Numbers of residues are shown in brackets. Those in unfavourable conformations (score~3.00) are labeled. Shading shows favourable conformations as obtained from an analysis of 163 structures at resolution 2.0 Å or better. (a) Arg (10), (b) Asn (6), (c) Asp (8), (d) Cys (5), (e) Gin (7), (f) Glu (10), (g) His (6), (h) Ile, (i) Leu (12), (j) Lys (11), (k) Met (2), (l) Phe (10), (m) Trp (2) and (n) Tyr (3).
Fig. 4: 3D structure of AC1 proteins of the two different begomovirus showing the diversity of the position of tyrosine

AC4 protein: ACMV-[CM] AC4 expressed in trans from the 35S promoter increased EACMCV DNA accumulation by ~8-fold compared to control transfection with EACMCV alone (Vanitharam et al., 2004). Mutation analysis of AC4 has shown that this protein is involved in symptoms development. Stanley and Latham (1992) introduced a stop codon at two different locations in the C4 gene without affecting the amino acid sequence of Rep. Agrobacterium inoculation of a mutant of TLCV (Australia) produced drastically reduced symptoms, although the level of viral DNA was similar to the wild type virus and suggests that C4 is not required by TLCV to replicate or to spread through the host plant but is involved in symptom development (Rigden et al., 1994).

The Ramachandran plot for AC4 of Abutilon mosaic virus (isolate West India) (AbMV) explains that out of the 1988 residues 89.4% residues in most favoured region and 10.6% residue in additional allowed regions (Fig. 5).
CONCLUSION

Homology modeling 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 five or ten best homologous proteins showing maximum sequence similarity with the unknown protein for homology search we may predict that their structure could be similar.

Breeding works towards the incorporation of begomovirus resistance genes must take into account the high degree of genetic diversity begomoviruses in the world. A given resistance gene might be extremely effective against a particular begomovirus species and totally ineffectual against distinct, unrelated species. Also, the rate of evolution of begomoviruses seems to be inordinately fast (Padidam et al., 1999), leading to the quick emergence of new strains or species that might overcome resistance genes. The durability of a resistance based on a single gene is therefore questionable. The genetic diversity of begomoviruses present in additional regions should be determined and the prevalence of distinct species or strains identified, in order to orient breeding programs towards resistance to the species prevalent in each region.

REFERENCES


