

RESEARCH ARTICLE

Structural insights into the RNA interaction with Yam bean Mosaic virus (coat protein) from *Pachyrhizus erosus* using bioinformatics approach

Varsha Acharya¹, R. Arutselvan¹, Kalidas Pati^{1*}, Ajaya Kumar Rout², Budheswar Dehury^{2#}, V. B. S. Chauhan¹, M. Nedunchezhiyan¹

1 Regional Centre, ICAR-Central Tuber Crops Research Institute Bhubaneswar, Odisha, India, **2** ICAR-Central Inland Fisheries Research Institute, Barrackpore, Kolkata, West Bengal, India

Current Address: Present address: ICMR-Regional Medical Research Centre, Chandrasekharpur, Bhubaneswar, Odisha, India

* kalidas9555@gmail.com



OPEN ACCESS

Citation: Acharya V, Arutselvan R, Pati K, Rout AK, Dehury B, Chauhan VBS, et al. (2022) Structural insights into the RNA interaction with Yam bean Mosaic virus (coat protein) from *Pachyrhizus erosus* using bioinformatics approach. PLoS ONE 17(7): e0270534. <https://doi.org/10.1371/journal.pone.0270534>

Editor: Rajarshi Gaur, Deen Dayal Upadhyaya Gorakhpur University, INDIA

Received: February 10, 2022

Accepted: June 11, 2022

Published: July 22, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0270534>

Copyright: © 2022 Acharya et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and [Supporting Information](#) files, including NCBI accession numbers.

Abstract

Plants are constantly threatened by a virus infection, i.e., Potyviruses, the second largest genus of plant viruses which results in several million-dollar losses in various essential crops globally. Yam bean (*Pachyrhizus erosus*) is considered to be one of the essential tuberous legume crops holding a great potential source of starch. Yam Bean Mosaic Virus (YBMV) of Potyvirus group belonging to the family potyviridae affects Yam bean and several angiosperms both in the tropical and sub-tropical regions causing large economical losses in crops. In this study, we attempted to understand the sequence-structure relationship and mode of RNA binding mechanism in YBMV CP using in silico integrative modeling and all-atoms molecular dynamics (MD) simulations. The assembly of coat protein (CP) subunits from YBMV and the plausible mode of RNA binding were compared with the experimental structure of CP from Watermelon mosaic virus potyvirus (5ODV). The transmembrane helix region is present in the YBMV CP sequence ranging from 76 to 91 amino acids. Like the close structural-homolog, 24 CPs monomeric sub-units formed YBMV a conserved fold. Our computational study showed that ARG¹²⁴, ARG¹⁵⁵, and TYR¹⁵¹ orient towards the inner side of the virion, while, THR¹²², GLN¹²⁵, SER⁹², ASP⁹⁴ reside towards the outer side of the virion. Despite sharing very low sequence similarity with CPs from other plant viruses, the strongly conserved residues Ser, Arg, and Asp within the RNA binding pocket of YBMV CP indicate the presence of a highly conserved RNA binding site in CPs from different families. Using several bioinformatics tools and comprehensive analysis from MD simulation, our study has provided novel insights into the RNA binding mechanism in YBMV CP. Thus, we anticipate that our findings from this study will be useful for the development of new therapeutic agents against the pathogen, paving the way for researchers to better control this destructive plant virus.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist

Introduction

Yam bean (*Pachyrhizus erosus*), also known as “Jicama” belongs to the family Fabaceae is an economically high-yielding leguminous tuber crop showing the widest adaptation with an enhanced nutritional value [1]. It is cultivated broadly in Mexico, Singapore, China, the Philippines, Indonesia, Hawaii, and India. It is the native crop of Tropical America, some parts of Africa, and Asia. In India, it is cultivated in many regions of Assam, Orissa, Bihar, Tripura and West Bengal [2]. Yam bean mosaic virus was discovered for the first time in Brazil [3]. Yam bean (*Pachyrhizus erosus*) is also one of the well-known legumes known as “Bengkluang” widely found in Indonesia. It is cultivated mainly in several regions of West and Central Java in Indonesia [4]. Infected plants typically exhibit leaf green vein banding, moulting, curling, inter-veinal mosaic, stunted development and the disease has been known to cause severe production losses in yam [5, 6]. The virus is distributed non-persistently through aphids and vegetative propagation. YMV has been identified in various Dioscoreaceae plants [7–10] and can be mechanically transferred to *Chenopodium amanticolor*, *Nicotiana megalosiphon* and *Nicotiana benthamiana* [8]. Yam bean serves a key function in delivering food and protein-rich micronutrients and protein for human use [10]. The subtribe Glycininae belongs to the genus *Pachyrhizus* which consists of an essential legume crop named soybean (*Glycine max*) [11, 12]. The genus involves three cultivated species *i.e.*, *Pachyrhizus tuberosus*, *Pachyrhizus erosus*, and *Pachyrhizus ahipa* produce roots with high storage capacity as well as with high proteins *i.e.* up to 12% in dry matter, iron content (up to 52 ppm) followed by zinc (up to 14 ppm) [13]. The genomic size of the yam bean is estimated to be between 572 and 597Mbp [14]. The most widely consumed yam bean species are *tuberosus* and *erosus*, which have different morphologies within the species and genus [15]. Viruses, on the other hand, are becoming a serious restriction for tuber crops and roots, particularly in the tropics [13].

Sorensen and co-authors reported that 20–40% of yield declination is due to certain viral diseases in *Pachyrhizus erosus* [13]. Despite of the negative impacts on yam bean crops, there is very little information available on viral diseases. The Yam bean (*P. erosus*) usually gets infected by bacteria like *Pseudomonas* [16], fungi [17, 18], insect-pests (*i.e.*, aphids, mealybug, and bruchids) [13], nematodes and viruses. It has been earlier reported that there are two major viruses destroying yam bean cultivation *i.e.*, bean common mosaic virus (BCMV) and sincama mosaic virus (SMV) [13]. Damayanti and colleagues identified a strain causing massive damage to Yam Bean (*Pachyrhizus erosus*) cultivation in different parts of Indonesia and recognized it as BCMV. Latter the virus was observed to be a distinct strain and named as BCMV-IYbn. Soon after in Peru, certain unique strains were observed that were suspected to be BCMV. Following that, a virus with a distinct strain was obtained for observation, and tests were performed to identify it [19]. Small interfering RNA (siRNA) sequencing took place and assembly was performed on a sample resulting in severe mosaic and leaf deformation [20]. Finally, after the analysis, it was proved that the virus was not BCMV and later Fuentes and co-authors in 2012 proposed the name of the new virus as Yam bean Mosaic Virus (YBMV). The YBMV CP comes under the Potyviridae family consisting of virions *i.e.*, non-enveloped as well as flexuous and filamentous. On the other hand, certain phylogenetic analyses were performed using only the CP (coat protein) of YBMV from the isolates of Vietnam [21] and Indonesia [19] and found maximum identity to YBMV over the CP encoding region which assures the presence of YBMV in the virus as identified as BCMV-IYbn previously. In Peru, Indonesia, and Vietnam, the presence of YBMV and associated viruses in a yard-long bean, yam bean, and black bean respectively suggests the presence of this virus in geographically remote areas and various crops [22]. The potyviral CP is a multifunctional protein that plays a part in almost every phase of the virus life cycle, in addition to its structural purpose of preserving the viral

DNA. Interactions between the viral RNA and CP, in particular, must be closely regulated for the RNA to be assigned to each of its roles correctly during the infection process [23].

Cucumber mosaic virus, infection causes mosaic symptoms in a wide range of agricultural plants, including cucumbers. The replacement of amino acid 129 in the coat protein of the Cucumber mosaic virus results in structural alterations in the virus [24]. The screening of potential inhibitors against the coat protein of the Apple Chlorotic Leaf Spot Virus was carried out. Apple chlorotic leaf spot virus (ACLSV) coat protein (CP) is an important latent virus on the apple plant that has been identified [25]. The coat protein (CP) of the tobacco mosaic virus (TMV) self-assembles in transgenic plants that have been deprived of viral RNA, forming aggregates that are dependent on the physical parameters of the environment [26]. A biosimulation technique was used to identify plant-based analogs that might serve as potential inhibitors against the tobacco mosaic virus [27]. siRNAs created by the plant's defensive system, the host RNA silencing mechanism, target the viral RNA or DNA and function in post-transcriptional gene silencing (PTGS) or transcriptional gene silencing (TGS) pathways, respectively, to inhibit viral replication. Host gene expression modulation by virus- and virus-derived small RNAs: new insights into the pathogenesis of viruses and viruses-derived small RNAs [28]. Because of its critical function in viral replication, the RNA-dependent RNA polymerase (RdRp) has emerged as a potential therapeutic target [29].

Previously, there was no protein-related information in the yam bean crop and thus there has been no work done on functional and structural protein characterization of YBMV globally. In 1988, Shukla and Ward investigated the structural characterization of the potyvirus (coat proteins) CP and observed that the amino acid sequences of CP are sufficient to identify and differentiate the individual potyvirus and its strains [30–32]. The coat protein is considered to be the major gene product in the virion [33]. The genome of potyvirus consists of eight different gene products of which the CP and the genome-linked protein are the solitary gene products in the virion [33].

Thus, in our study, we have taken the coat protein of YBMV for the structural and functional analysis of the virus CP. Since genomic structure differs significantly among families and the CPs display very low-level sequence similarity, it is fascinating to explore the virion architecture. Through integrative modeling, we explored the plausible organization of YBMV CP and its mode of single-stranded RNA (ssRNA) recognition followed by the all-atoms molecular dynamics simulation (MDS) for a 50 ns timeframe. We also identified a few conserved residues that aid in binding of RNA which suggests antiviral compounds could be targeted to these regions which may profound effect on the genome assembly and packaging of economically important viruses like YBMV. As a consequence of our research, we now have a better understanding of the structural basis of CP assembly in YBMV, which will help us in developing new antiviral drugs that target the conserved fold of RNA binding.

Materials and methods

Primary sequence analysis

The FASTA format amino acid sequence of YBMV CP was obtained from the NCBI protein database with accession number YP_004940328. The coat protein functional region/domain region was identified using the Conserved Domain Database (CDD) [34]. The phosphorylation site was identified using the NetPhos 2.0 server [35]. To gain broad chemistry of YBMV CP, a clear study on the primary sequence was carried out using the ProtParam tool [36] of ExPASy Proteomic Server. The ProtParam provides information on different Physico-chemical properties of the coat protein, such as instability index, molecular weight, iso-electric point, grand average hydropathy (GRAVY), and aliphatic index.

Molecular evolutionary analysis of YBMV CP

The molecular evolutionary analysis of YBMV, a phylogenetic tree was built using the NJ technique in MEGA version 6 software [37]. A total of twenty-six homologous coat protein sequences of potyvirus were collected using the DELTA-BLASTP search against the non-redundant (nr) database to construct the evolutionary genetic tree. Percentage of trees replicated where the accompanied taxa grouped in the bootstrap test (containing 1000 replicates) were signified next to the branches. The position correction method is used for determining the protein evolutionary distances through gap extension penalty, gap separation distance, and gap opening penalty employing Blossum weight matrix. All locations encompassing gaps and missing data were removed throughout the study. Using Clustal W software, multiple sequence alignment was conducted on fourteen coat proteins of Potyviruses [38].

Secondary structural analysis

The PSIPRED server was used to identify the secondary structure of YBMV CP [39]. PSIPRED helps in sequence annotation by TM (transmembrane) domain mapping and secondary structure mapping [40]. This server is one of the specific secondary structure prediction methods that includes two feed-forward neural networks in which the analysis is based on the output obtained from Delta-Blastp.

Template identification

Suitable templates were detected for the comparative modeling of YBMV CP, the target sequences of YBMV CP were investigated for similar sequences using the Blast-P *i.e.*, against the PDB database [41]. The matrix used for the template identification is the Blossum-62 matrix with a default inclusion threshold value (0.005) and E-value (10). The templates were chosen based on the sequence identity, less E-value, query coverage, and structural resolution. After the selection of an appropriate template, a target-template alignment was performed for the generation of protein structure.

Computational modeling

Arnold and co-authors have developed the SWISS-MODEL workspace with a range of tools that permits the user to modify and validate the various modeling steps manually. The SWISS-MODEL workstation includes application software required for 3D structure in an easy-to-use web-based modeling computer unit. A 3D model for the target protein is generated based on the sequence alignment among the target and template structures. For structure creation, SWISS-MODEL was used to model the 3D proteins [42].

Model validation

An essential web application such as PROCHECK has been deployed to explain the authenticity, stereochemical consistency, and durability of the projected 3D model [43]. PROCHECK was used for accessing the geometry of the model. Protein structure analysis *i.e* ProSA [44] was applied for calculating the energy potential of the predicted model, and the Verify3D [45] score used for the analysis to check the quality of the predicted model, It's also used to check if the atomic model is compatible with its amino acid sequence. Mol Probitly [46] is a web service for structure validation providing a broad-spectrum solidly based evaluation of model quality at both the local and global levels for both nucleic acids and proteins [46]. In comparison to other approaches that are optimized to find native structures, ERRAT [47] was performed for estimating the accuracy of the non-bonded atoms of the modeled YMBV. The validation

results were interpreted, accompanied by re-validation to achieve an initial optimized model structure, and the structural ambiguities were resolved.

Identification of RNA binding pocket

To identify the plausible mode of RNA recognition, the homo-oligomeric assembly (24-mers) [48] using monomeric sub-units of modeled CP of YBMV was built. Based on the potyvirus virion structure (PDB: 5DOV) [49], the oligomeric structure of CP sub-units was modeled and transferred the coordinates of RNA to our modeled structure using PyMOL. The electrostatic surface potential map was constructed using the APBS plug-in of PyMOL. To identify, the conserved RNA binding site which was reported in earlier studies [48]. We have employed HADDOCK v2.4 web server to predict the RNA binding affinity with both mutant and wild-type [50–52].

Molecular dynamics simulation

The stability and dynamic behavior of the modeled structure were investigated based upon the Desmond with 2005 force field i.e., (OPLS) Optimized Potentials for Liquid Simulations [53]. To examine the multiple conformations and interactions of the ligands within the binding pocket of YBMV CP, we used Desmond molecular dynamics simulations. TIP3P was used as a solvent in an orthorhombic box and the protein was neutralized by introducing counter ions and 0.15M of salt concentration to neutralize the protein [54]. The distance between the box wall and the protein was fixed at 10 Å microns to avoid steric interference with the protein periodic images. Energy minimization was performed using a hybrid approach of steepest descent and limited memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithms with a maximum of 5000 steps until a gradient threshold was reached up to 25 kcal/mol [55]. The NVT (number of particles, Volume, and Temperature) ensemble conditions were kept the same for 1ns to get simulation data for post-simulation analysis [56]. Nose-Hoover thermostats were used to keep the temperature at 300 degrees Celsius throughout the simulation, and the Martyna-Tobias-Klein barostat method was employed to keep the pressure at a consistent level throughout. A multi-step RESPA integrator technique was utilized to investigate the equation of motion in dynamics [57]. The completed equilibrated system was used to run a 50 ns MD simulation, which was examined using Desmond's event analysis module. The docking complexes were employed as starting structures for a 50 ns MD simulation, which was performed on them. Structures were imported into Desmond's set-up wizard and soaked with TIP3P inside an orthorhombic box, like a box with a top and bottom. All other procedures were carried out in the same manner as described in the free form of the YBMV CP protein simulation.

Results

Sequence analysis of YBMV CP

Yam bean mosaic virus (YBMV) coat protein (CP) sequence was retrieved from the NCBI protein database (YP_004940328) consisting of 275aa. The domain prediction using SMART, Pfam, and CDD (conserved domain database) revealed a common conserved region of the coat protein sequence of the YBMV i.e. (41-270aa) 230aa. This conserved region is utilized further for the structural analysis. The primary sequence analysis of YBMV CP (230aa long) illustrated that the protein is acidic (an isoelectric point is 6.26) in nature and has a molecular mass of 26.41kDa. The aliphatic index is 75.13 which indicate the protein's stability over a wide range of temperature. It is widely recognized that a protein with an instability index less than 40 is

considered to be stable, while if the value ranges above 40 it shows the protein to be unstable [58]. The coat protein (CP) of YBMV is stable as its instability index is 37.02 (*i.e.*, <40). The protein contains 31 negatively charged and 29 positively charged amino acids. YBMV CP has a Grand Average Hydropathicity GRAVY value of -0.576, indicating that it may have better contact with water. The phosphorylation sites of YBMV CP protein were determined using Net-Phos 2.0 [35] which indicates 5 serine, 5 threonine, and 2 tyrosine sites for phosphorylation.

Secondary structure analysis and 3D prediction

The secondary sequence analysis reveals that, sequence annotation is influenced by expected features like TM and secondary structure mappings. PSIPRED performs the secondary structure analysis where the coils/turns get dominated over strands and helices along with the secondary structure elements. The secondary structure map discovers the amino acid residues which have the propensity to be a beta-sheet or an alpha-helix are present in the proteins derived from the YBMV CP sequence. In addition, sequence characteristics such as the DomPred boundary, disordered area, disordered protein binding sites, and DomSSEA boundary are discovered and annotated in the sequence (S1 Fig). The protein sequence of YBMV CP shows the presence of the transmembrane helix region starting from 76aa up to 91aa. It also depicts the disordered protein binding area, which spans 215 to 233 amino acids. Sequence features like cytoplasmic region, pore-lining helix and extracellular region are annotated and predicted. The contact maps of YBMV CP revealed the amino acid residue-residue interactions (S2 Fig).

Phylogenetic analysis

To understand the molecular evolution of YBMV from *Pachyrhizus erosus*, phylogenetic analysis was performed on 26 different coat proteins of Potyviruses belonging to the family Potyviridae (see S1 Table). The Neighbor-Joining method with 1000 iterations was used to create the 2D phylogenetic tree and branch length 4.229 has been displayed using MEGA version 6 (Fig 1). The Poisson correction technique was used to compute the evolutionary distances. The phylogenetic cladogram represents altogether five clusters and three major clusters formation. On the other hand, it was found that the CP sequence of YBMV shows closely related to CP of Soybean mosaic virus, Bean common mosaic virus and Watermelon mosaic virus. Clusters 3, 4 and 5 show highly dissimilar clades *i.e.*, less sequence similarity to YBMV coat protein. Multiple sequence alignment was performed using Clustal W with the coat protein sequences of Yam bean mosaic virus, Dasheen mosaic virus, Leek yellow stripe potyvirus, Chilli veinal mottle virus, Watermelon mosaic virus, Beet mosaic virus, Onion yellow dwarf virus, Bean common mosaic virus, Cocksfoot streak virus, Bean yellow mosaic virus, Cowpea aphid-borne mosaic virus, Johnson grass mosaic virus, Soybean Mosaic virus, Lily mottle virus, Maize dwarf mosaic virus, Papaya ringspot virus, Peanut mottle virus, Yam mosaic virus, Japanese yam mosaic virus, Zucchini yellow mosaic virus, Turnip mosaic virus, Wild potato mosaic virus, Sugarcane mosaic virus, Sweet potato feathery mottle virus, Sorghum mosaic virus. Multiple sequence alignment of YBMV CP (coat protein) sequence with other CP viruses from the potyvirus family (S3 Fig).

3D modeling

The search for the templates for YBMV CP in PDB using Delta Blastp explained the crystal structure of Chain A, Watermelon Mosaic Virus (WMV) coat protein (PDB ID: 5ODV), Chain: A [48] shared the highest homology with an identity with coat protein (Bean common mosaic virus) of 80.44% and Query coverage 97%; Coat protein (Peanut stripe virus) identity 79.57 and Query coverage of 97%. The target-template alignment was conducted using the ESPript 3.0 programme as the comparative modelling hinges on a sequence alignment

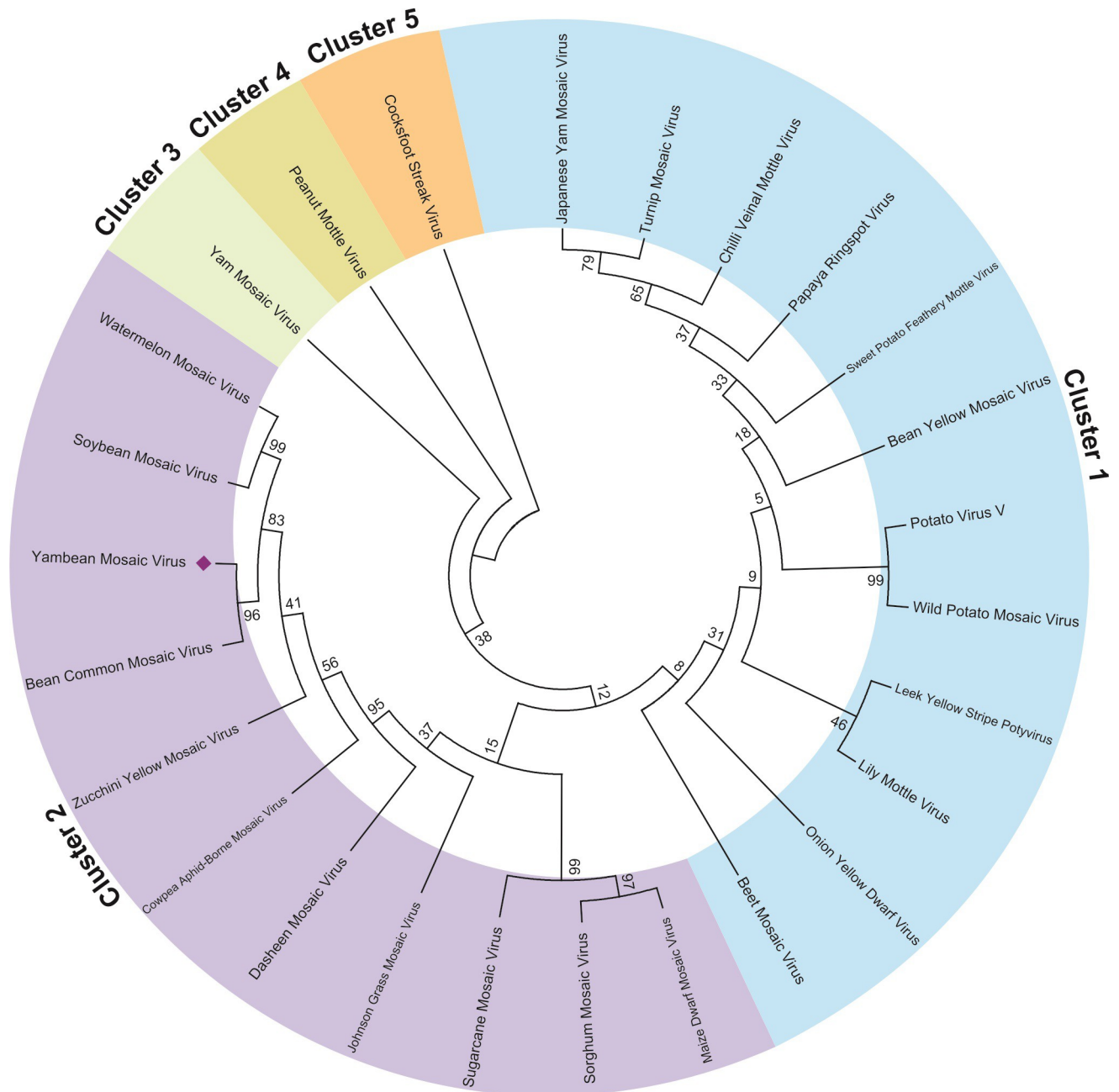


Fig 1. The evolutionary relationship of taxa of YBMV CP and other closely related organisms using the NJ method. The bootstrap test (1000 repetitions) showed the percentage of replicate trees where the related taxa were grouped. The analysis was conducted using MEGA version 6.0 software.

<https://doi.org/10.1371/journal.pone.0270534.g001>

between the template and the target sequence with an experimentally determined structure (Fig 2). Based on the target-template alignment, the protein model was constructed by the SWISS-MODEL web server.

Structure validation

The protein structure of the YBMV coat protein comprises two β -strands (one parallel and one anti-parallel) and eight α -helices. The outer surface of the protein is surrounded by the α -

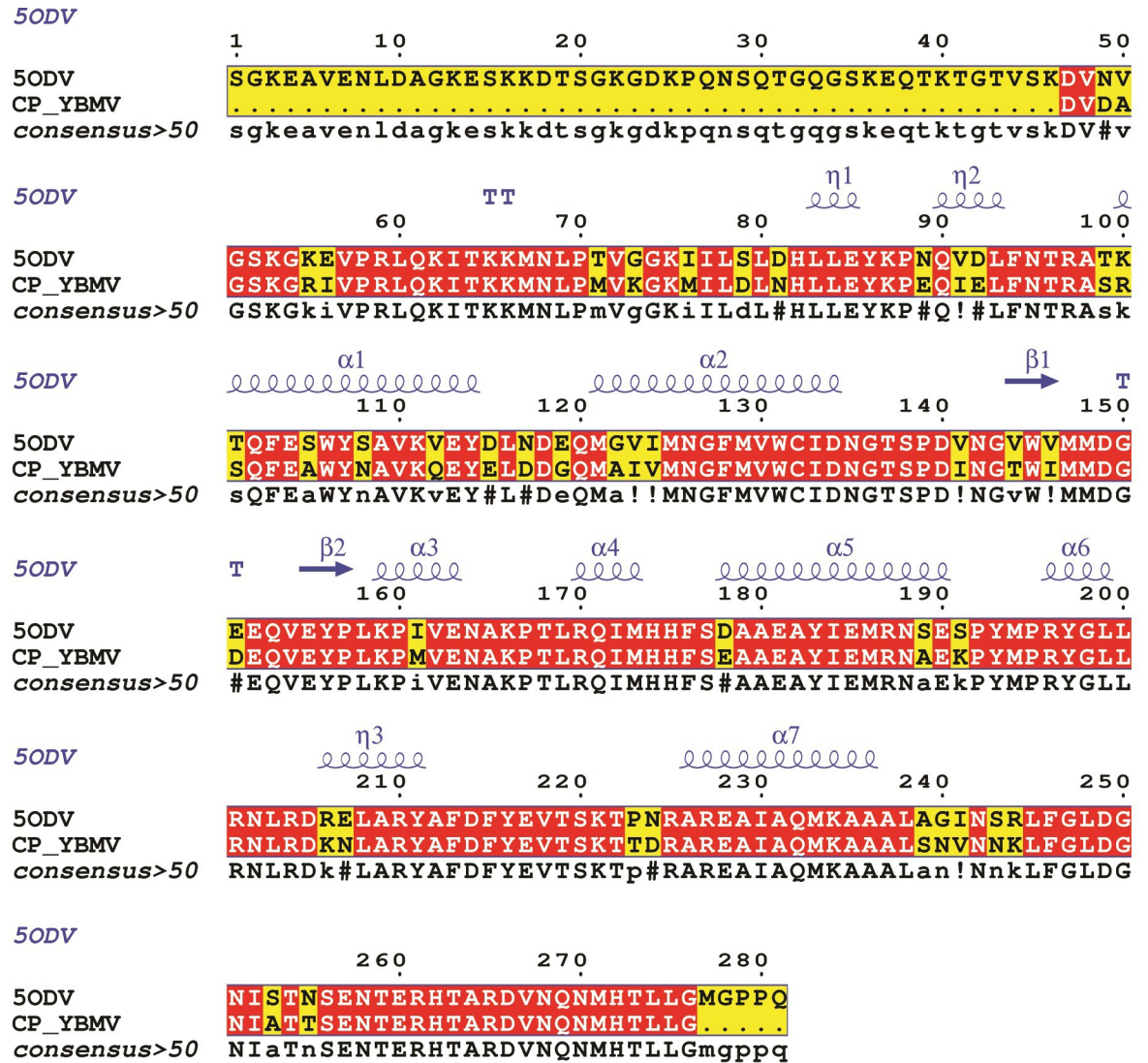


Fig 2. Target-template alignment of YMBV CP (coat protein) sequence with 5ODV using ESPript 3.0 software.

<https://doi.org/10.1371/journal.pone.0270534.g002>

helices and the inner core of the model is built of β -sheets (Fig 3A and 3B). The dihedral angles (Φ/Ψ) of the YBMV coat protein's accuracy were measured by using the Ramachandran plot in Procheck program (Fig 4A). Secondary structure analysis was performed using the PROCHECK analysis tool thus representing the alpha helices and beta sheets in a distinct manner. The 3D model represented a good percentage of residues in favored regions (86.2%), 13.5% residues in additional allowed regions, 0% residues in disallowed regions and 0.5% residues in allowed regions. ERRAT is used in the statistical calculation of the interactions between different atoms types, it also reveals a good validation score is 90%. ERRAT suggests the reliability of the initial model. The ProSA software aims to obtain the proposed model's energy profile as well as the z-score value, which further helps to measure the energy of residue relationship. ProSA evaluated a z-score rating value *i.e.*, as depicted in Fig 4B. Protein topology map, constructed using Procheck, with beta-strands as a pink-colored arrow and the alpha-helices as red cylinders (S4 Fig). The details of the protein structure validation are summarized in Table 1.

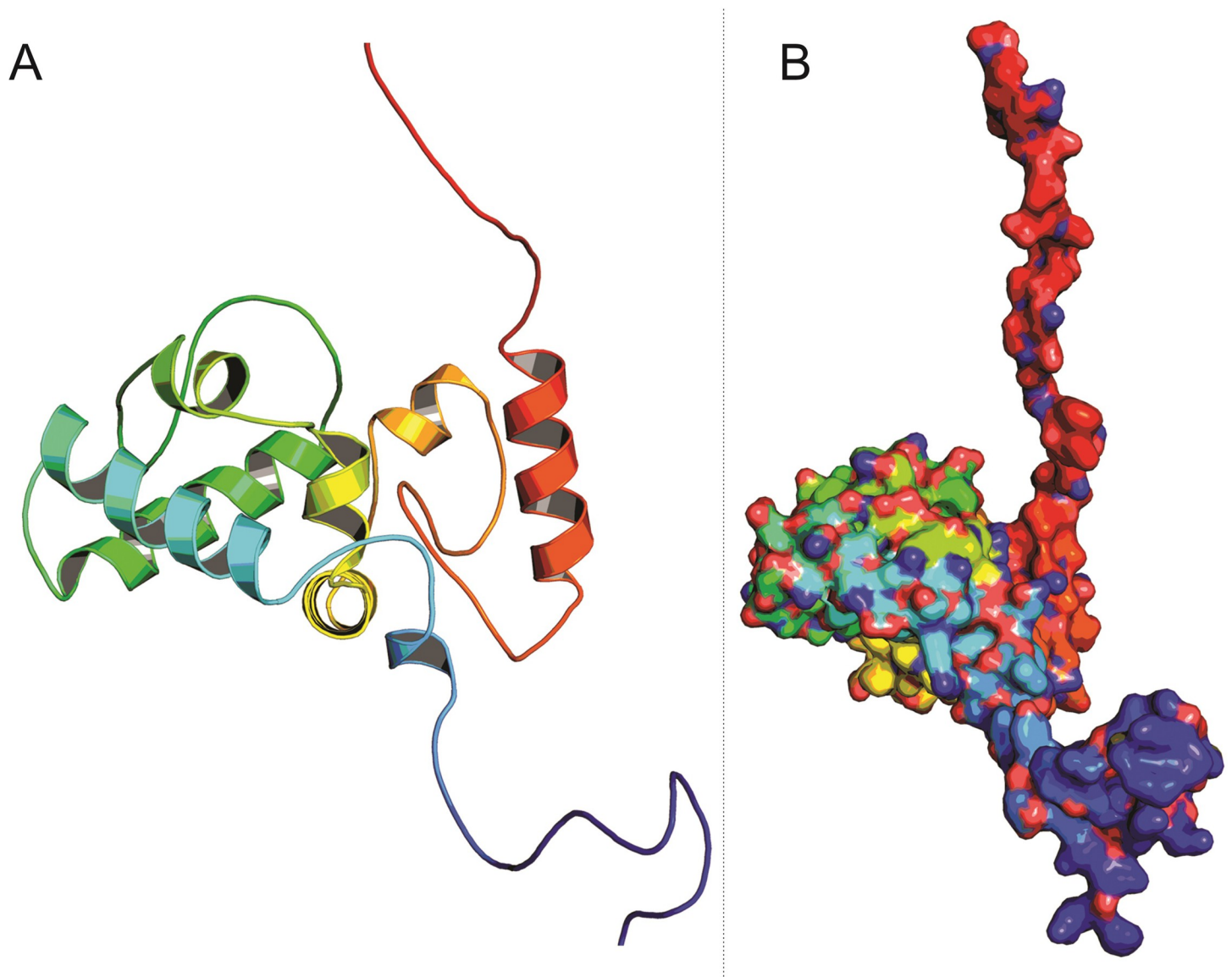


Fig 3. (A) The three-dimensional structure of YBMV CP N-terminal region (red); C-terminal region (blue). (B) The protein is visualized as a solid ribbon representing β -sheets, α -helices, and coils.

<https://doi.org/10.1371/journal.pone.0270534.g003>

Oligomeric assembly of the conserved fold of YBMV coat proteins

It has been observed that 24 CPs monomeric sub-units form a conserved fold. As a result, we produced its oligomeric form based on the template in our research. The method [53] is based on supervised machine learning and Supports Vector Machines, which integrate interface conservation with other template characteristics structural clustering to produce a quaternary structure quality evaluation. The QSQE score is a number between 0 and 1 that indicates the predicted accuracy of interchain interactions for a model generated with a certain alignment and template. The greater numbers imply more reliance. This is added to the GMQE score, which measures the correctness of the resultant model's tertiary structure.

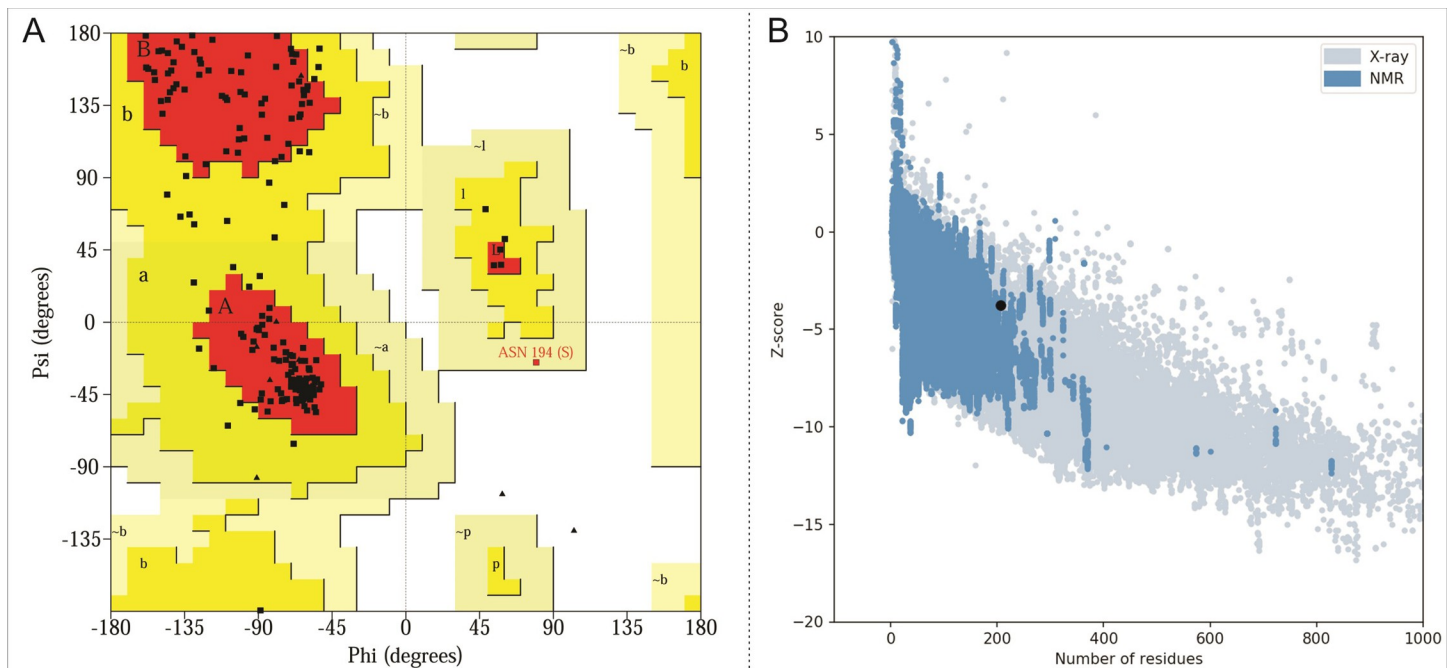


Fig 4. (A) Ramachandran plot of the YBMV CP model. The plot was created using the PROCHECK program. (B) Protein Structure Analysis (ProSA) of model YBMV CP. The overall quality of the YBMV CP model shows the z-score value of -3.75 (Native conformation to its template).

<https://doi.org/10.1371/journal.pone.0270534.g004>

RNA binding sites

Interactions between the CP and the viral RNA, in particular, must be tightly controlled for the RNA to be correctly assigned to each of its functions throughout the infection process. In the case of YBMV CP, each subunit covers nucleotides of ssRNA (Fig 5A–5C). The presence of a conserved binding site is suggested by analyzing the seven amino acid residues present in the atomic structure of the YBMV CP. The key residues interacting with wild type RNA binding sites are LYS 188, VAL 172, GLU 171, ASN 197, ARG 124 and THR 122 respectively (Fig 6A).

Table 1. Model validation scores of yam bean mosaic virus coat protein (YBMV CP).

Model Validation Tools	Model validation reports	Scores
PROCHECK	Most favored region (%)	86.2%
	Additional allowed region (%)	13.3%
	Generously allowed region (%)	0.5%
	Disallowed region (%)	0.0%
	Overall G-factor	0.18
Verify3D	Averaged 3D-1D Score > 0.2	61.35%
ERRAT	Overall quality (%)	90.00
ProSA	Z-score	-3.75
ProQ	Predicted LG score	8.547
	Predicted Max Sub	-0.257
Molprobrity	Ramachandran outliers (%)	0.00%
	Poor rotamers (%)	0.00%
	Bad bonds (%)	0.29%
	Bad angles (%)	1.34%

<https://doi.org/10.1371/journal.pone.0270534.t001>

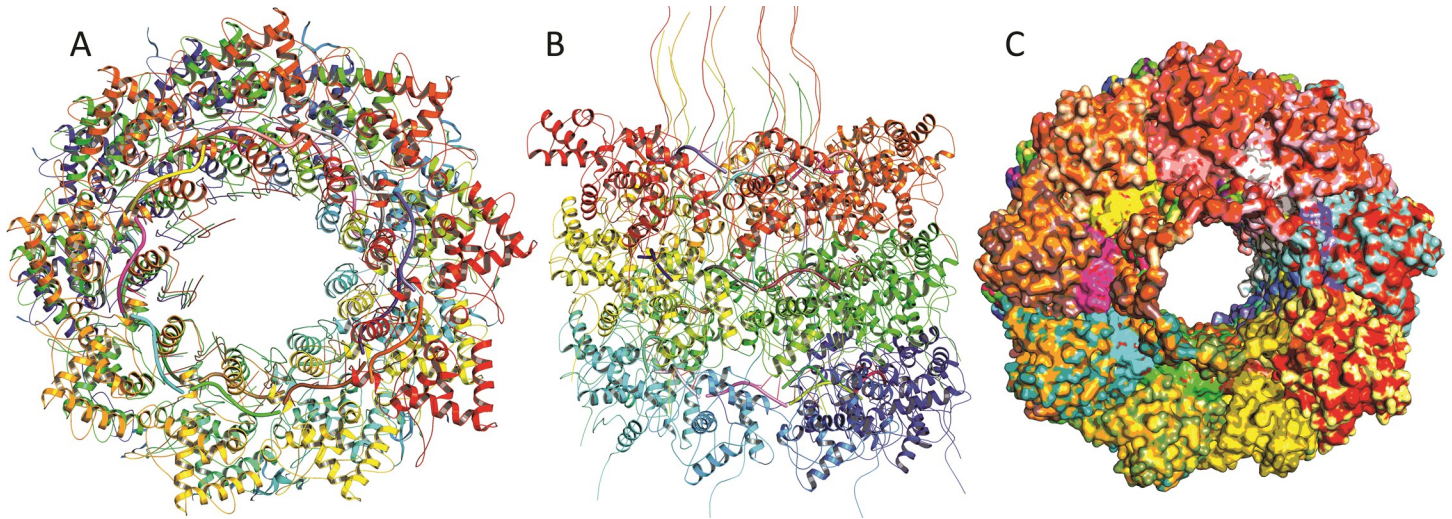


Fig 5. Oligomeric assembly of 24 identical subunits of CP with RNA (A) top, (B) front and (C) surface representation).

<https://doi.org/10.1371/journal.pone.0270534.g005>

The key residues interacting with mutant type RNA binding sites are LYS 188, SER 92, THR 91, PRO 93, ASP 168, TYR 170, ALA 142, LEU 123 and THR 122 respectively (**Fig 6B**). The intermolecular contacts of YBMV CP with RNA were also determined using SWISS-MODEL shown in **Table 2**.

A total of 132 structures in 11 cluster(s), which represents 66% of the water-refined models HADDOCK generated. The wild type HADDOCK score (-68.7), Van der Waals energy (-38.6), Electrostatic energy (-218.7) respectively. The total of 160 structures in 8 cluster(s), which represents 80% of the water-refined models HADDOCK generated. The mutant type HADDOCK score (-61.9), Van der Waals energy (-40.6), Electrostatic energy (-166.2) respectively.

Molecular dynamics simulation of YBMV CP

Several dynamic stability parameters, including the RMSD, $C\alpha$ -RMSF, Rg of intermolecular H-bonds, were examined with simulation time after the protein and complex of YBMV CP were obtained. RMSD evaluated the initial structure by MD simulation for 50 ns to investigate the effective confirmation of hits, protein stability, and RNA-binding recognition. By analyzing the deviation caused during the simulation, the Root-mean-square deviation (RMSD) was able to estimate the protein's stability in comparison to its initial structure. The RMSD value shows that, the RNA complex has a more stable shape. The average RMSD of the complex and 3D model of YBMV CP was computed 50 ns after the start of the simulation. RMSD for complex 30 ns to 50 ns stable within 4Å while apo protein deviated throughout the simulation from 6–7.5 (**Fig 7A**). YBMV CP was produced in both the apo and holo forms, and the C-alpha atom's RMSF profile for each residue was computed to determine the residue-wise mobility during the simulation. The YBMV CP-RNA system showed more flexibility as compared to the YBMV CP. The complex's C-alpha Root mean square fluctuations (RMSFs) demonstrated that high fluctuation and confirmed the RMSD results. The mean variation plot was used to assess the compounds average RMSF value. The RMSF for the protein complex C-alpha displays less variation than the RMSF for the apo protein (**Fig 7B**). From the examination of the radius of gyration (Rg), it is clear that the compactness of the complex protein was stabilized between 17.0–17.5 Å, but the compactness of the apo protein was disturbed

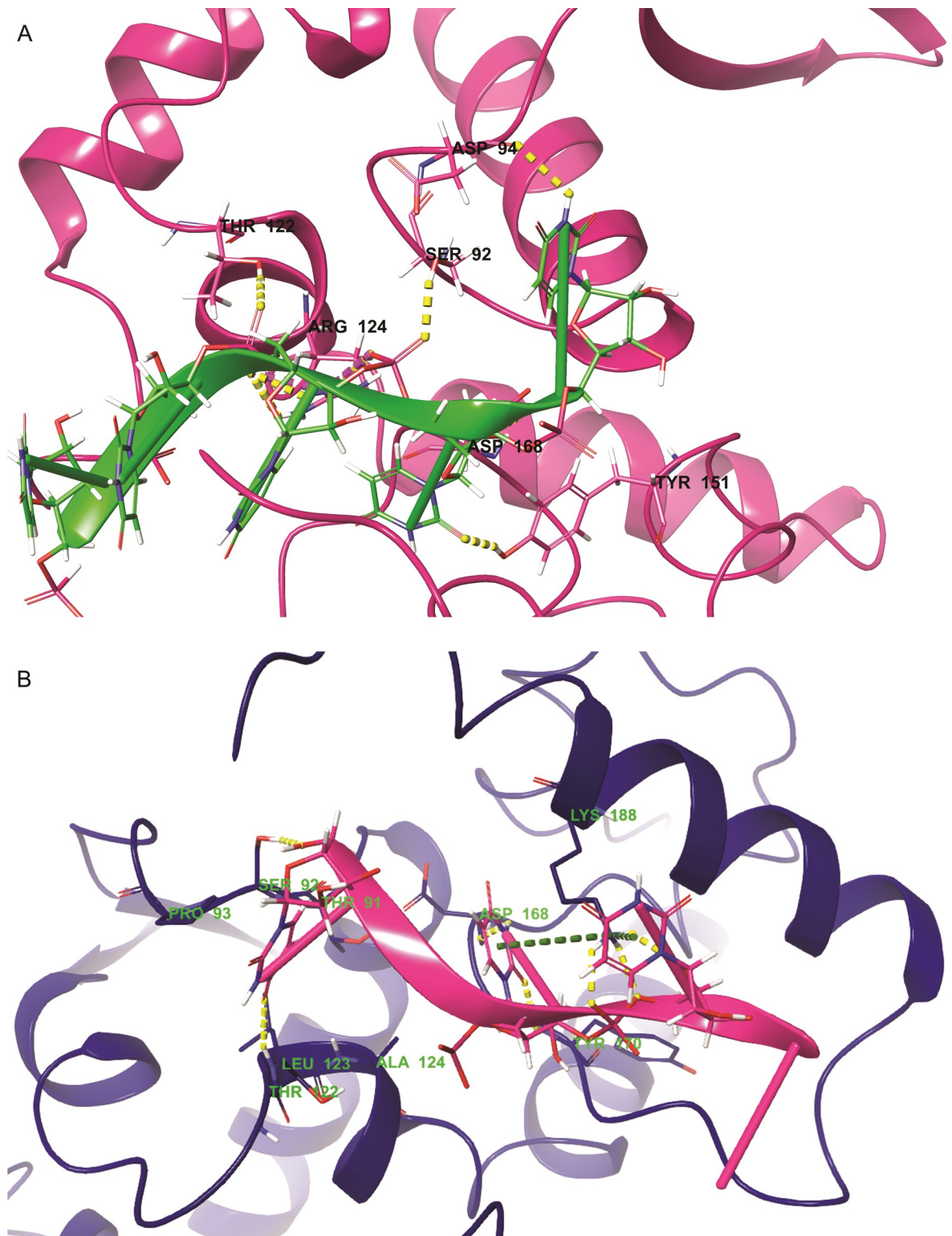


Fig 6. Interaction analysis of YBMV CP with RNA binding reorganization. (A) Wild type; (B) R124A mutant.

<https://doi.org/10.1371/journal.pone.0270534.g006>

throughout the simulation. Protein and RNA complexes' compactness and dimensions are determined by it, and this may explain why they interact well. Protein and RNA complexes

Table 2. Intermolecular contacts YBMV CP with RNA.

Pairs	Distance	Type	Category
A:ARG124:NH2—B:U3:OP2	2.17	Hydrogen Bond;Electrostatic	Salt Bridge
A:ARG124:NH2—B:U4:OP2	3.57	Hydrogen Bond;Electrostatic	Salt Bridge
A:ARG155:NH2—B:U5:OP1	3.04	Hydrogen Bond;Electrostatic	Salt Bridge
A:ARG124:NH1—B:U2:OP1	5.55	Electrostatic	Attractive Charge
A:ARG124:NH1—B:U3:OP1	5.59	Electrostatic	Attractive Charge
A:SER92:OG—B:U4:OP1	2.04	Hydrogen Bond	Conventional Hydrogen Bond
A:THR122:OG1—B:U3:OP1	3.08	Hydrogen Bond	Conventional Hydrogen Bond
A:TYR151:OH—B:U4:O2	2.65	Hydrogen Bond	Conventional Hydrogen Bond
B:U5:H3—A:ASP94:O	2.43	Hydrogen Bond	Conventional Hydrogen Bond
B:U3—A:LEU192	5.26	Hydrophobic	Pi-Alkyl
B:U4—A:LYS188	5.23	Hydrophobic	Pi-Alkyl

<https://doi.org/10.1371/journal.pone.0270534.t002>

compactness and dimensions are determined by it, and this may explain why they interact well (Fig 7C). H-bonds are the most essential factors for biomolecular complexes' stability. It was determined that the number of discrete H-bonds formed between certain amino acid residues of the YBMV core protein and RNA atoms was computed using the Desmond with Optimized Potentials for Liquid Simulations (OPLS) 2005 force field. We computed the intermolecular H-bonds generated between YBMV CP and the RNA-Complex as a function of the simulation timeframe. The H-bond study revealed that RNA created 8–12 hydrogen bonds with protein during the replication process (Fig 7D). Overall, taking into consideration of dynamics stability, H-bond analyses and the pre and post-MD RNA binding views, it may be concluded that was more stable.

Wild type

The ARG side chain has four structure-based donors that can all hydrogen-bond to a single H-bond acceptor.

Asp 168 showed more than 95% direct interaction and formed water mediated interaction more than 35% with the RNA during simulation. Arg124 established 3 direct interactions with the RNA and stabilized more than 60% of the molecular dynamics simulation (Fig 8A). Fig 8B represent the different bond interaction of the amino acid residues with RNA. Arg124 interacted with RNA by forming mostly H-bond and water bridge. ARG 155 communicated with RNA through H-bond, salt bridge and water bridge. The bottom panel shows which residues interact with the ligand in each trajectory frame. Some residues make more than one specific contact with the ligand, which is represented by a darker shade of orange, according to the scale to the right of the plot (Fig 8C).

Mutant type MD simulation

We have used different stability predictors to assess the effect of mutation of the key residues. The RMSD value shows that the R124A and complex has a more stable during 50 ns MD simulation. RMSD for complex 15 ns to 45 ns stable within 4Å while apo protein deviated throughout the simulation (Fig 9A). The complex's C-alpha Root mean square fluctuations (RMSFs) demonstrated that high fluctuation and confirmed the RMSD results. The RMSF for the R124A and complex C-alpha displays less variation than the RMSF for the apo protein (Fig 9B). The examination of the radius of gyration (Rg), it is clear that the compactness of the complex protein was stabilized between 17.0–15.5 Å, but the

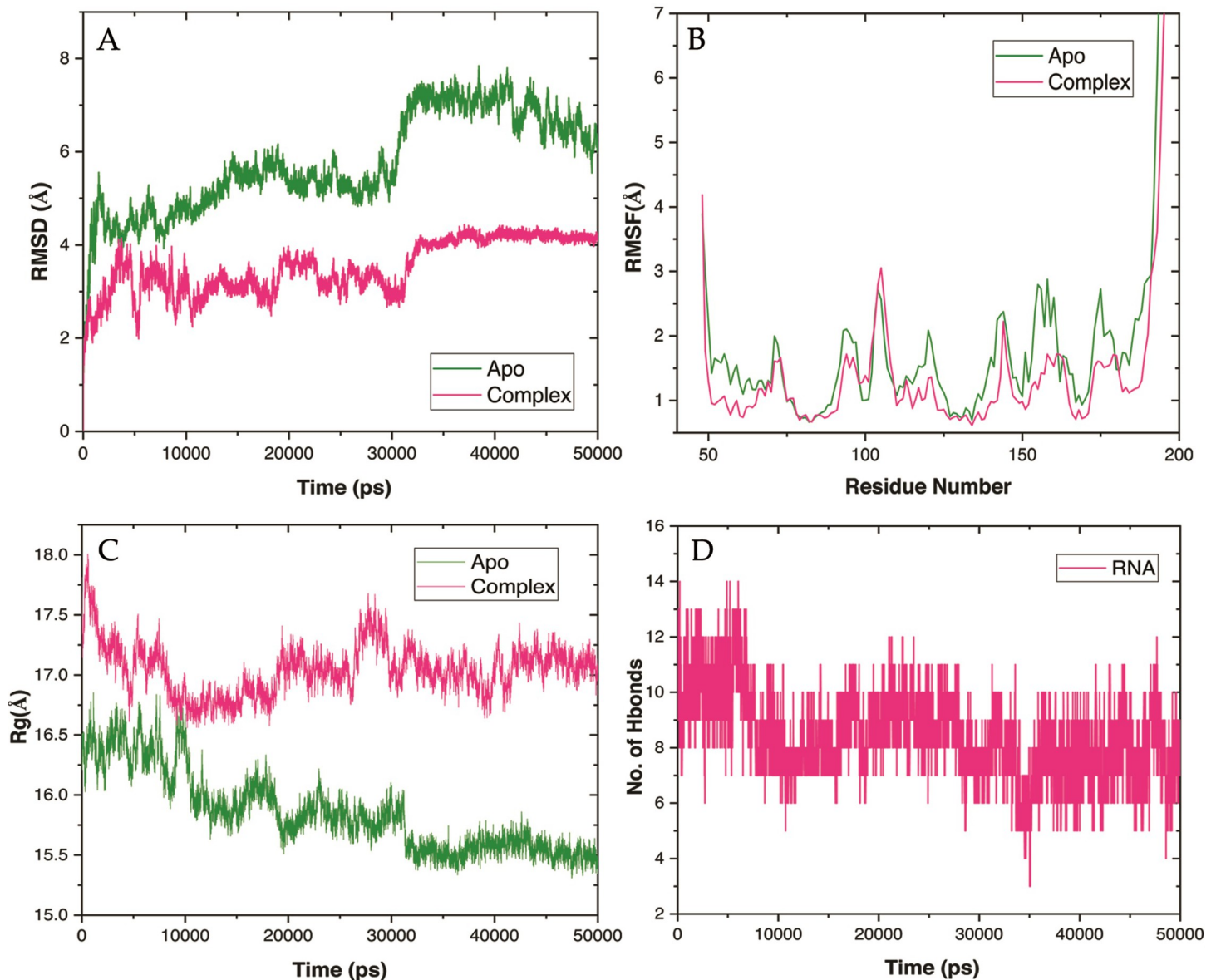


Fig 7. Conformational stability of YBMV CP of apo and complex through MD simulation. (A) Backbone RMSD (B) C-alpha RMSF YBMV CP (C) Radiation of gyration for YBMV CP; (D) Hydrogen bond analysis over 50 ns period concerning the index of the residue.

<https://doi.org/10.1371/journal.pone.0270534.g007>

compactness of the R124A protein was disturbed throughout the simulation (Fig 9C). H-bonds are the most essential factors for biomolecular complexes' stability. It was determined that the number of discrete H-bonds formed between certain amino acid residues of the R124A and R124A Complex was computed using the Desmond with Optimized Potentials for Liquid Simulations (OPLS) 2005 force field. We computed the intermolecular H-bonds generated between R124A and R124A Complex as a function of the simulation timeframe. The H-bond study revealed that created 2–14 hydrogen bonds with protein during the replication process (Fig 9D). Overall, taking into consideration of dynamics stability, H-bond analyses and the pre and post-MD RNA binding views, it may be concluded that was more stable.

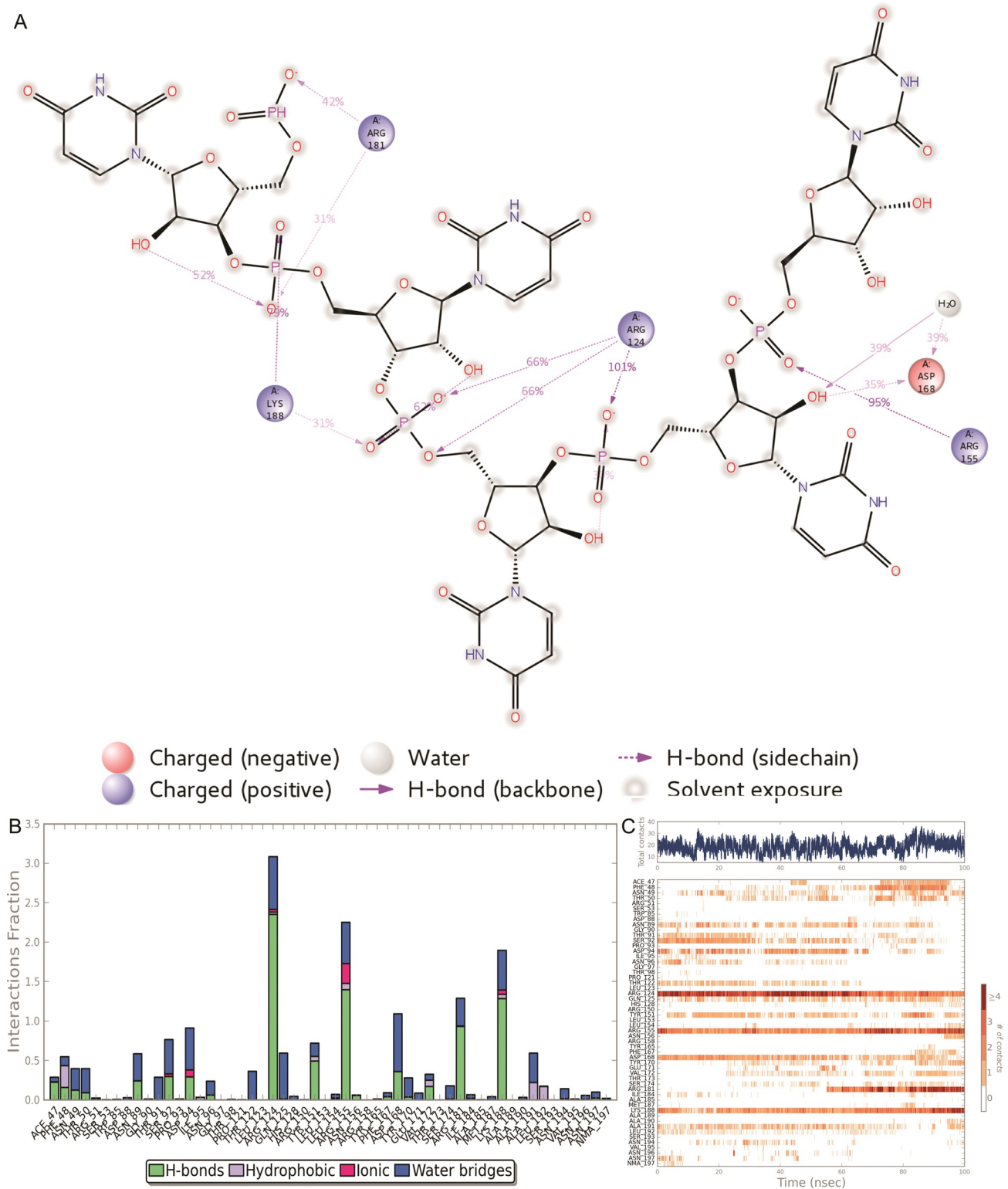


Fig 8. The H-bond occupancy of wild type YBMV CP with RNA throughout the trajectory. (A) 2D interaction of RNA with YBMV CP during simulation (B) the amino acid residues involved in the interactions (C) the overall interaction of the protein with RNA.

<https://doi.org/10.1371/journal.pone.0270534.g008>

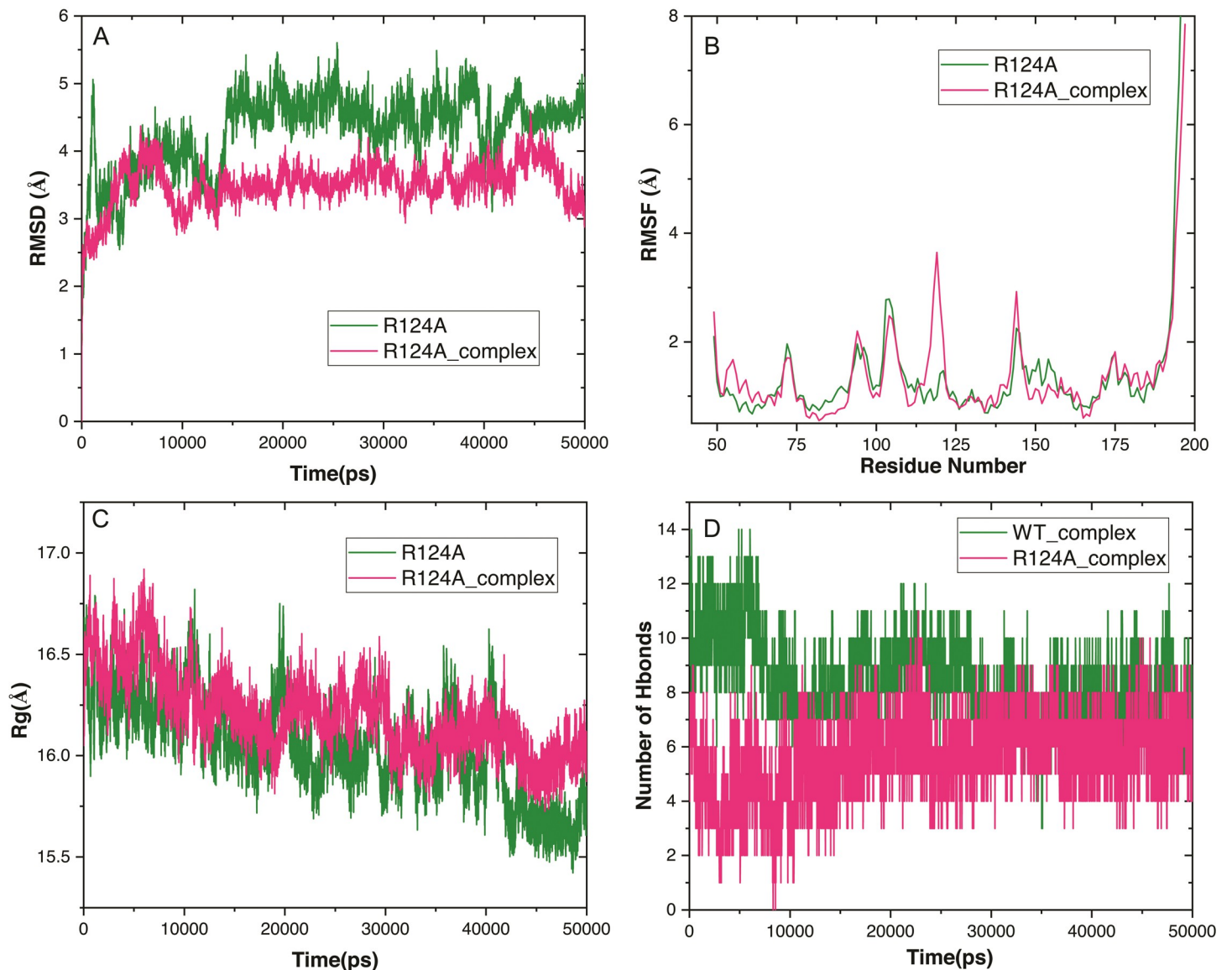


Fig 9. Conformational stability of YBMV CP of R124A_Apo and R124_Complex through MD simulation. (A) Backbone RMSD (B) C-alpha RMSF YBMV CP (C) Radiation of gyration for YBMV CP; (D) comparison of Hydrogen bond analysis between wild type and R124A mutant over 50 ns period concerning the index of the residue.

<https://doi.org/10.1371/journal.pone.0270534.g009>

Mutant type

The Fig 10A disclosed that Ala124, Leu123, Pro121, Tyr170, Lys180 and Thr176 established direct interaction with the RNA and stabilized more than 30% throughout simulation. Additionally, it has been noticed that ARG155 communicated with RNA through H₂O molecule. Furthermore, distribution of interaction of each amino acid residues has been displaced in Fig 10B. A timeline representation of the interactions and contacts H-bonds, Hydrophobic, Ionic, Water bridges. The bottom panel shows which residues interact with the RNA in each trajectory frame. Some residues make more than one specific contact with the ligand, which is represented by a darker shade of orange, according to the scale to the right of the plot (Fig 10C).

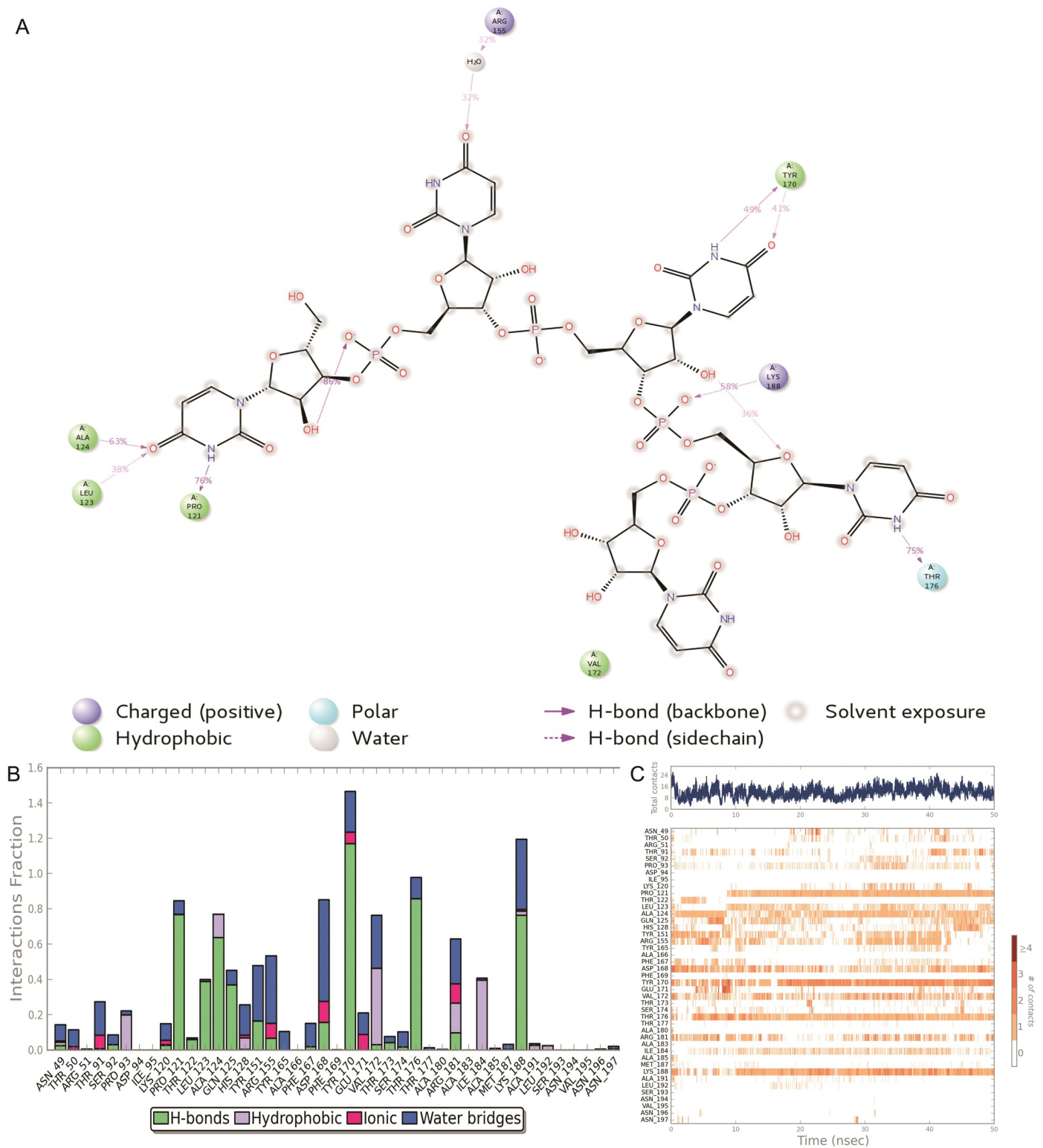


Fig 10. The H-bond occupancy of mutant type YBMV CP with RNA throughout the trajectory. (A) 2D interaction of RNA with YBMV CP during simulation (B) the amino acid residues involved in the interactions (C) the overall interaction of the protein with RNA.

<https://doi.org/10.1371/journal.pone.0270534.g010>

Discussion

Among legumes, yam beans (*Pachyrhizus erosus*) have the greatest adaptability and production potential [10]. Potyvirus is considered to be the largest and economically most important plant virus. The potyvirus consists of 28 plant virus families and groups. In 1974 total of 1112 species of 369 genera in 53 plant families has been reported [30]. One-quarter of known plant RNA viruses belong to the potyvirus group family potyviridae [59, 60]. Presently there are 2,026 no. of plant species and 566 no. of genera along with 81 botanical families that are globally distributed, which are subjected to potyviruses [61]. More than 200 species of aphids are responsible for the transmission of the potyvirus occurring in a non-persistent way [59, 62]. As a result, potyviruses have a unique capacity to adapt to novel habitats, hosts, and vectors, according to Nigam and co-authors. The investigation of polyprotein-wide and genome-wide variation indicated that the potyviral genome possesses hypervariable locations that preferentially collect nucleotide substitutions and sites under positive selection, encouraging protein disorder and perhaps determining host adaptability [63]. The potyviruses are found in numerous pathogens or strains which vary in biological properties *i.e* pathogenicity [64]. According to Fuentes et al., (2012), a new virus known as Yam Bean Mosaic Virus (YBMV) was discovered and named, and it belongs to the potyviridae family. The virus was first reported by [19], and by default, the name was given as BCMV-IYbn. Latter in 2012 after further study on the virus Fuentes discovered the virus in the yam bean crop and named it Yam Bean Mosaic Virus (YBMV). There is no clear identity, of the YBMV CP core amino acid sequence with the corresponding sequences of members of the genus Potyvirus. It is necessary to achieve a better knowledge of the coat protein structure of YBMV which may help to understand the virus's multiple functions. The N-terminal portions of potyvirus CP sequences are the most diverse, whereas the C-terminal half is very conserved [62]. Thus, the capsid protein and 3' NCR sequences are the most essential parts of the potyviral genome for taxonomic study, and the 3'-terminal area may be utilized to identify virus strains from other potyviruses, according to all molecular data [62, 65].

According to Shukla and Ward (1989), the nucleotide sequence of the CP gene among strains of a certain potyvirus species was more than 90% similar [31], whereas the degree of similarity in the nucleotide sequences of the 3' UTR across strains of specific potyvirus species ranged from 83–99 percent, whereas different virus species had identities ranging from 39–53 percent, according to Frenkel et al (1989) [66]. Adams et al 2005 looked at several potyvirus sequences and determined that the greatest possible CP nucleotide sequence identity in the same potyvirus species was 76–77 percent, while the highest possible CP amino acid sequence identity was 82 percent [67].

The virus is described as a new, positive-sense, single-stranded RNA virus enveloped in a novel way [19]. Instead of the full-length sequence, only the domain region was taken for the structural analysis for accurate optimization. The viral sequence is depicted as a novel enveloped, negative-sense as well as single-stranded RNA virus whose strain belongs to the country named Peru. The virus comes under the potyvirus group, a family named potyviridae. The overall small RNA size distribution revealed a significant preponderance of 21-nt RNAs, which is consistent with prior severe potyviral infections [20]. A phylogenetic analysis was carried out representing five clusters from which two are considered to be major clusters. From our study, it was found that the YBMV CP shows high similarity with Bean Common Mosaic Virus (BCMV) and then secondly with Soybean Mosaic Virus and then with Watermelon Mosaic Virus, the remaining cluster illustrates highly dissimilar clades. Currently, the isolate taken for the analysis is from Peru [22] and the respective amino acid sequence is available in NCBI. The primary structure analysis, as well as the sequence annotation, is essential for

converting the sequence data to functional genes which is essential to understanding the functional and biological mechanisms [68]. In addition to direct functional analysis, the precise position of the transmembrane helices (TM) is important for functional annotation [69]. The YBMV contains the TM region starting from 76aa up to 91aa. For understanding, the functional and the biological mechanisms sequence annotation and physiochemical property analysis are essential, which are obtained after converting the sequence data into functional genes [70]. Preliminary attempts were made to obtain structural information on the coat protein of YBMV by using *in-silico* approaches yielding certain results.

It is important to know how proteins work in three dimensions to answer many biological questions. However, the number of genes and genomes that have been sequenced is quickly outpacing the number of experimentally determined structures. In this approach, comparative modeling continues to be a highly effective strategy of choosing. It contributes to the closure of the gap between what we know about the sequence of a protein and how it appears by creating precise and trustworthy protein models [71].

To solve many biological problems there is a need to have good knowledge regarding the three-dimensional protein structure of the virus (YBMV). Nowadays, comparative modeling is performed as it is an increasingly important method to obtain a suitable 3D protein structure [70, 72–75]. Comparative modeling's main benefit is that it aids in filling in gaps between available sequence and structural data by generating a reliable and accurate protein model [71]. The approach and parameters of MD simulation were adapted from our early studies to better comprehend the molecular function of unknown/hypothetical molecules [76–81].

In our work, 3D modeling is performed following the template search for YBMV coat protein (PDB ID: 5ODV, Chain: A) [48], resulting in appropriate 3D models for a wide range of applications. However, the model, structural stability, and protein-ligand interaction studies also provide prior knowledge for treatments creation soon. Molecular dynamics simulation provides a very realistic, safe, and straightforward atomic level technique for understanding protein folding and dynamics over a certain timescale [82].

Using a bioinformatics approach, several computational tools are provided to examine various dynamical behaviours of the protein structure. In this work, we examined our protein and the complex of YBMV CP utilizing several characteristics such as RMSD, CRMSF, and Rg of intermolecular H-bonds. The original structure was analyzed for RMSD using MD simulation for 50ns to study the effective confirmation of hits, protein stability, and RNA-binding recognition. The average root mean square deviation (RMSD) of the complex and three-dimensional model of YBMV CP was calculated 50ns after the simulation. The apo protein's RMSD fluctuated from 6–7.5 throughout the experiment, while the RMSD for the complex remained steady at 4. (Fig 7A). Less variation is seen in the RMSF for C- α , a complex of proteins, than in the RMSF for apo C- α . (Fig 7B). The radius of gyration (Rg) shows that the compactness of the complex protein was stable between 17.0–17.5, but the compactness of the apo protein was not stable at all (Fig 7C). H-bonds are the most important variables in determining the stability of biomolecular complexes. A force field based on the Desmond with Optimized Potentials for Liquid Simulations (OPLS) 2005 force field was used to calculate the number of discrete H-bonds generated between specific amino acid residues of the YBMV core protein and RNA atoms. According to this research, we computed the intermolecular H-bonds produced between YBMV-CP and the complex of RNA molecules as a function of the simulation period. It was discovered through the H-bond study that RNA formed 8–12 hydrogen bonds with protein during the replication process (Fig 7D).

The emergence of this virus disease might pose a significant threat to India's newly launched and promising yam bean Mosaic Virus output. Other significant factors to assess the threat of this virus to global yam bean output include the virus's worldwide prevalence,

variability, susceptibility, and influence on the yield of different yam bean species. Our results will also help in the development of novel antiviral medicines that target the RNA binding pocket, which might disrupt the genome packaging and assembly of economically significant plant viruses like YBMV [48]. Overall, we anticipate that our unique insights into RNA interaction mechanism in Yam bean Mosaic Virus (coat protein), as well as the new study on virus particles, will lead to control of potyviral infection in the future.

Conclusions

Understanding the significance of a viral protein in the infection process involves a detailed three-dimensional structure of the protein. Our study substantiates from the molecular evolutionary analysis that the YBMV CP of yam bean is highly similar to the coat protein of Bean common mosaic virus, Soybean mosaic virus, Watermelon mosaic virus and potyviridae family. We systematically searched the desired YBMV CP sequence for the *in-silico* analysis and homology-driven structure prediction was generated and characterized in yam bean. The emergence of this new viral illness might pose a significant danger to the yam bean production that has only lately begun and appears to be promising. The entire genome of a novel potyvirus could be reconstructed from the siRNA sequences, and it was distinct enough from other sequences to be classified as a new species. Future research should focus on developing precise detection methods and determining the degree of seed transfer so that suitable phytosanitary measures may be developed and implemented to prevent the disease from spreading further. Other critical characteristics needed to assess the threat of this virus to global yam bean output include the virus's worldwide prevalence, variability, susceptibility, and influence on the yield of diverse yam bean species.

Supporting information

S1 Fig. The secondary structure map. The figure represents the secondary structure map where query sequence and coloured residues are presented as per the annotations made by each analysis methods. Map shows helical residues that are coloured in pink, b-strand residues coloured in yellow, putative domain boundaries are indicated in blue, and green bordered box represents disordered protein binding.

(TIF)

S2 Fig. The contact map. The figure shows the contact map of YBMV protein model. The map shows a matrix with contacts with the color scale which represents the relationship between the amino acids in the particular map.

(TIF)

S3 Fig. Multiple sequence alignment. Multiple sequence alignment of YBMV_CP (coat protein) sequence with other CP viruses from the potyvirus family. The image is generated in Multalin software.

(TIF)

S4 Fig. Protein topology map. Protein topology map, constructed using Procheck, with beta-strands as a pink-coloured arrow and the alpha-helices as red cylinders.

(TIF)

S1 Table. Identification of protein domain. The full-length sequence acc. no. of the coat protein of potyvirus and identification of protein domain using the conserved domain search.

(DOC)

Acknowledgments

The authors are thankful to the Director, ICAR-Central Tuber Crops Research Institute, Trivandrum, Kerala, India for kind support to carry out the research work.

Author Contributions

Conceptualization: R. Arutselvan, Kalidas Pati.

Data curation: Ajaya Kumar Rout, Budheswar Dehury.

Formal analysis: Varsha Acharya, Ajaya Kumar Rout, Budheswar Dehury.

Investigation: Varsha Acharya, Kalidas Pati.

Methodology: Varsha Acharya.

Software: Varsha Acharya, Ajaya Kumar Rout, Budheswar Dehury.

Supervision: R. Arutselvan, Kalidas Pati.

Validation: Kalidas Pati.

Writing – original draft: Varsha Acharya, R. Arutselvan, Ajaya Kumar Rout, Budheswar Dehury.

Writing – review & editing: Ajaya Kumar Rout, Budheswar Dehury, V. B. S. Chauhan, M. Nedunchezhiyan.

References

1. Reddy PP. Plant protection in tropical root and tuber crops. Springer; 2015.
2. Pati K, Behera S, Ray DK, Behera PR, Sarkar A, Chauhan VBS, et al. Element Profiling of Thirty Genotypes of Yam bean in eastern India by Using Proton Induced X-ray Emission (PIXE). *J Food Compos Anal.* 2021; 104006.
3. de Sá Andrade Medeiros L, de Oliveira IA, Pires LL, Kitajima EW, da Silva GS, Beserra JEA. First report of Yambean mosaic virus in Brazil. *Australas Plant Dis Notes.* 2019; 14: 1–4.
4. Lim TK. *Pachyrhizus erosus*. Edible medicinal and non-medicinal plants. Springer; 2016. pp. 465–481.
5. Thouvenel J-C, Dumont R. Perte de rendement de l'igname infectée par le virus de la mosaïque en Côte d'Ivoire. 1990.
6. Adeniji MO, Shoyinka SA, Ikotun T, Asiedu R, Hughes J, Odu BO. Yield loss in Guinea yam (*Dioscorea rotundata* Poir.) due to infection by Yam mosaic virus (YMV) genus Potyvirus. *Ife J Sci.* 2012; 14: 237–244.
7. Thouvenel J, Fauquet C. Yam mosaic, a new potyvirus infecting *Dioscorea cayenensis* in the Ivory Coast. *Ann Appl Biol.* 1979; 93: 279–283.
8. Odu BO, Hughes J d'A, Asiedu R, Ng NQ, Shoyinka SA, Oladiran OA. Responses of white yam (*Dioscorea rotundata*) cultivars to inoculation with three viruses. *Plant Pathol.* 2004; 53: 141–147.
9. Ita EE, Uyoh EA, Nakamura I, Ntui VO. Efficient elimination of Yam mosaic virus (YMV) from white yam (*Dioscorea rotundata* Poir.) by cryotherapy of axillary buds. *South African J Bot.* 2020; 130: 123–129. <https://doi.org/10.1016/j.sajb.2019.12.022> PMID: 32565604
10. National Academy of Sciences (Washington D. Tropical Legumes: Resources for the Future: Report of an Ad Hoc Panel of the Advisory Committee on Technology Innovation, Board on Science and Technology for International Development, Commission on International Relations, National Research Council. National Academy of Sciences; 1979.
11. Tay Fernandez CG, Pati K, Severn-Ellis AA, Batley J, Edwards D. Studying the genetic diversity of yam bean using a new draft genome assembly. *Agronomy.* 2021; 11: 953.
12. Lee J, Hymowitz T. A molecular phylogenetic study of the subtribe Glycininae (Leguminosae) derived from the chloroplast DNA rps16 intron sequences. *Am J Bot.* 2001; 88: 2064–2073. PMID: 21669638
13. Sørensen M. Yam Bean: *Pachyrhizus* DC.-Promoting the conservation and use of underutilized and neglected crops. 2. *Biodiversity International*; 1996.

14. Pati K, Zhang F, Batley J. First report of genome size and ploidy of the underutilized leguminous tuber crop Yam Bean (*Pachyrhizus erosus* and *P. tuberosus*) by flow cytometry. *Plant Genet Resour.* 2019; 17: 456–459.
15. Sørensen M, Døygaard S, Estrella JE, Kvist LP, Nielsen PE. Status of the South American tuberous legume *Pachyrhizus tuberosus* (Lam.) Spreng.: field observations, taxonomic analysis, linguistic studies and agronomic data on the diversity of the South American *Pachyrhizus tuberosus* (Lam.) Spreng. complex with speci. *Biodivers Conserv.* 1997; 6: 1581–1625.
16. Birch RG, Alvarez AM, Patil SS. A bacterial leaf spot caused in yam bean by *Pseudomonas syringae* pv. *phaseolicola*. *Phytopathology.* 1981; 71: 1289–1293.
17. Annerose DJM, Diouf O. The Yam Bean Project at Bambey, rapport annuel 1993–1994. Second Annu Prog Report, STD3 Contract No ERBTS3* CT920115 (M Sørensen, ed). 1995; 137–151.
18. Díaz Arguelles A. El cultivo de la jícama en el Estado de Guanajuato. Instituto Nacional de Investigaciones Agrícolas, El Bajío (México). Centro . . .; 1979.
19. Damayanti TA, Susilo D, Nurlaelah S, Sartiami D, Okuno T, Mise K. First report of Bean common mosaic virus in yam bean [*Pachyrhizus erosus* (L.) Urban] in Indonesia. *J Gen Plant Pathol.* 2008; 74: 438–442.
20. Kreuze JF, Perez A, Untiveros M, Quispe D, Fuentes S, Barker I, et al. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology.* 2009; 388: 1–7. <https://doi.org/10.1016/j.virol.2009.03.024> PMID: 19394993
21. Ha C, Reville P, Harding RM, Vu M, Dale JL. Identification and sequence analysis of potyviruses infecting crops in Vietnam. *Arch Virol.* 2008; 153: 45–60. <https://doi.org/10.1007/s00705-007-1067-1> PMID: 17906829
22. Fuentes S, Heider B, Tasso RC, Romero E, Zum Felde T, Kreuze JF. Complete genome sequence of a potyvirus infecting yam beans (*Pachyrhizus* spp.) in Peru. *Arch Virol.* 2012; 157: 773–776. <https://doi.org/10.1007/s00705-011-1214-6> PMID: 22227800
23. Martínez-Turiño S, García JA. Potyviral coat protein and genomic RNA: A striking partnership leading virion assembly and more. *Virus Assem Exit Pathways.* 2020; 108: 165. <https://doi.org/10.1016/bs.avir.2020.09.001> PMID: 33837716
24. Bhardwaj VK, Purohit R. Structural changes induced by substitution of amino acid 129 in the coat protein of Cucumber mosaic virus. *Genomics.* 2020; 112: 3729–3738. <https://doi.org/10.1016/j.ygeno.2020.04.023> PMID: 32353478
25. Purohit R, Kumar S, Hallan V. Screening of potential inhibitor against coat protein of apple chlorotic leaf spot virus. *Cell Biochem Biophys.* 2018; 76: 273–278. <https://doi.org/10.1007/s12013-017-0836-z> PMID: 29380245
26. Sharma J, Purohit R, Hallan V. Conformational behavior of coat protein in plants and association with coat protein-mediated resistance against TMV. *Brazilian J Microbiol.* 2020; 51: 893–908. <https://doi.org/10.1007/s42770-020-00225-0> PMID: 31933177
27. Sharma J, Bhardwaj VK, Das P, Purohit R. Plant-based analogues identified as potential inhibitor against tobacco mosaic virus: A biosimulation approach. *Pestic Biochem Physiol.* 2021; 175: 104858. <https://doi.org/10.1016/j.pestbp.2021.104858> PMID: 33993976
28. Ramesh S V, Yogindran S, Gnanasekaran P, Chakraborty S, Winter S, Pappu HR. Virus and viroid-derived small RNAs as modulators of host gene expression: molecular insights into pathogenesis. *Front Microbiol.* 2021; 3170. <https://doi.org/10.3389/fmicb.2020.614231> PMID: 33584579
29. Singh R, Bhardwaj VK, Purohit R. Potential of turmeric-derived compounds against RNA-dependent RNA polymerase of SARS-CoV-2: An in-silico approach. *Comput Biol Med.* 2021; 139: 104965. <https://doi.org/10.1016/j.combiomed.2021.104965> PMID: 34717229
30. Shukla DD, Ward CW. Amino acid sequence homology of coat proteins as a basis for identification and classification of the potyvirus group. *J Gen Virol.* 1988; 69: 2703–2710.
31. Shukla DD, Ward CW. Identification and classification of potyviruses on the basis of coat protein sequence data and serology. *Arch Virol.* 1989; 106: 171–200. <https://doi.org/10.1007/BF01313952> PMID: 2673154
32. Shukla DD, Gough KH, Ward CW. Coat protein of potyviruses. *Arch Virol.* 1987; 96: 59–74.
33. Domier LL, Shaw JG, Rhoads RE. Potyviral proteins share amino acid sequence homology with picorna-, como-, and caulimoviral proteins. *Virology.* 1987; 158: 20–27. [https://doi.org/10.1016/0042-6822\(87\)90233-9](https://doi.org/10.1016/0042-6822(87)90233-9) PMID: 18644561
34. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 2017; 45: D200–D203. <https://doi.org/10.1093/nar/gkw1129> PMID: 27899674

35. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol.* 1999; 294: 1351–1362. <https://doi.org/10.1006/jmbi.1999.3310> PMID: 10600390
36. Garg VK, Avashthi H, Tiwari A, Jain PA, Ramkete PW, Kayastha AM, et al. MFPPI–multi FASTA Prot-Param interface. *Bioinformatics.* 2016; 12: 74. <https://doi.org/10.6026/97320630012074> PMID: 28104964
37. Tamura K, Stecher G, Peterson D, Filipksi A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013; 30: 2725–2729. <https://doi.org/10.1093/molbev/mst197> PMID: 24132122
38. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinforma.* 2003; 2–3.
39. Buchan DWA, Ward SM, Lobley AE, Nugent TCO, Bryson K, Jones DT. Protein annotation and modelling servers at University College London. *Nucleic Acids Res.* 2010; 38: W563–W568. <https://doi.org/10.1093/nar/gkq427> PMID: 20507913
40. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics.* 2000; 16: 404–405. <https://doi.org/10.1093/bioinformatics/16.4.404> PMID: 10869041
41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
42. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018; 46: W296–W303. <https://doi.org/10.1093/nar/gky427> PMID: 29788355
43. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 1993; 26: 283–291.
44. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 2007; 35: W407–W410. <https://doi.org/10.1093/nar/gkm290> PMID: 17517781
45. Eisenberg D, Lüthy R, Bowie JU. [20] VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.* 1997; 277: 396–404. [https://doi.org/10.1016/s0076-6879\(97\)77022-8](https://doi.org/10.1016/s0076-6879(97)77022-8) PMID: 9379925
46. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D Biol Crystallogr.* 2010; 66: 12–21. <https://doi.org/10.1107/S0907444909042073> PMID: 20057044
47. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* 1993; 2: 1511–1519. <https://doi.org/10.1002/pro.5560020916> PMID: 8401235
48. Zamora M, Méndez-López E, Agirrezabala X, Cuesta R, Lavín JL, Sánchez-Pina MA, et al. Potyvirus virion structure shows conserved protein fold and RNA binding site in ssRNA viruses. *Sci Adv.* 2017; 3: eaao2182. <https://doi.org/10.1126/sciadv.aao2182> PMID: 28948231
49. Khare D, Hale WA, Tripathi A, Gu L, Sherman DH, Gerwick WH, et al. Structural basis for cyclopropanation by a unique enoyl-acyl carrier protein reductase. *Structure.* 2015; 23: 2213–2223. <https://doi.org/10.1016/j.str.2015.09.013> PMID: 26526850
50. Van Zundert GCP, Rodrigues J, Trellet M, Schmitz C, Kastrius PL, Karaca E, et al. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J Mol Biol.* 2016; 428: 720–725. <https://doi.org/10.1016/j.jmb.2015.09.014> PMID: 26410586
51. Gnanasekaran P, Gupta N, Ponnusamy K, Chakraborty S. Geminivirus betasatellite-encoded β C1 protein exhibits novel ATP hydrolysis activity that influences its DNA-binding activity and viral pathogenesis. *J Virol.* 2021; 95: e00475–21.
52. Gnanasekaran P, Ponnusamy K, Chakraborty S. A geminivirus betasatellite encoded β C1 protein interacts with PsbP and subverts PsbP-mediated antiviral defence in plants. *Mol Plant Pathol.* 2019; 20: 943–960. <https://doi.org/10.1111/mpp.12804> PMID: 30985068
53. Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci Rep.* 2017; 7: 1–15.
54. Das M, Prakash S, Nayak C, Thangavel N, Singh SK, Manisankar P, et al. Dihydroactinidiolide, a natural product against A β 25–35 induced toxicity in Neuro2a cells: Synthesis, in silico and in vitro studies. *Bioorg Chem.* 2018; 81: 340–349. <https://doi.org/10.1016/j.bioorg.2018.08.037> PMID: 30189414
55. Vijayalakshmi P, Selvaraj C, Singh SK, Nisha J, Saipriya K, Daisy P. Exploration of the binding of DNA binding ligands to Staphylococcal DNA through QM/MM docking and molecular dynamics simulation. *J Biomol Struct Dyn.* 2013; 31: 561–571. <https://doi.org/10.1080/07391102.2012.706080> PMID: 22881193

56. Nayak C, Singh SK. In silico identification of natural product inhibitors against Octamer-binding transcription factor 4 (Oct4) to impede the mechanism of glioma stem cells. *PLoS One*. 2021; 16: e0255803. <https://doi.org/10.1371/journal.pone.0255803> PMID: 34613998
57. Patidar K, Deshmukh A, Bandaru S, Lakkaraju C, Girdhar A, Vr G, et al. Virtual screening approaches in identification of bioactive compounds Akin to delphinidin as potential HER2 inhibitors for the treatment of breast cancer. *Asian Pacific J Cancer Prev*. 2016; 17: 2291–2295. <https://doi.org/10.7314/apjcp.2016.17.4.2291> PMID: 27221932
58. Guruprasad K, Reddy BVB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng Des Sel*. 1990; 4: 155–161. <https://doi.org/10.1093/protein/4.2.155> PMID: 2075190
59. Gibbs AJ, Ohshima K, Phillips MJ, Gibbs MJ. The prehistory of potyviruses: their initial radiation was during the dawn of agriculture. *PLoS One*. 2008; 3: e2523. <https://doi.org/10.1371/journal.pone.0002523> PMID: 18575612
60. Gibbs A, Ohshima K. Potyviruses and the digital revolution. *Annu Rev Phytopathol*. 2010; 48: 205–223. <https://doi.org/10.1146/annurev-phyto-073009-114404> PMID: 20438367
61. Rivett DE, Ward CW, Belkin LM, Ramshaw JAM, Wilshire JFK. *The Lennox Legacy: The History of the CSIRO Laboratory at 343 Royal Parade Parkville*. CSIRO PUBLISHING; 1996.
62. Shukla DD, Ward CW, Brunt AA. *The Potyviridae*. CAB international; 1994.
63. Nigam D, LaTourrette K, Souza PFN, Garcia-Ruiz H. Genome-wide variation in potyviruses. *Front Plant Sci*. 2019; 10: 1439. <https://doi.org/10.3389/fpls.2019.01439> PMID: 31798606
64. Hollings M, Brunt AA. Potyvirus Group CMI/AAB Descrip. *Plant Viruses*. 1981;245.
65. Pappu SS, Brand R, Pappu HR, Rybicki EP, Gough KH, Frenkel MJ, et al. A polymerase chain reaction method adapted for selective amplification and cloning of 3' sequences of potyviral genomes: application to dasheen mosaic virus. *J Virol Methods*. 1993; 41: 9–20. [https://doi.org/10.1016/0166-0934\(93\)90158-n](https://doi.org/10.1016/0166-0934(93)90158-n) PMID: 8432765
66. Frenkel MJ, Ward CW, Shukla DD. The use of 3' non-coding nucleotide sequences in the taxonomy of potyviruses: application to watermelon mosaic virus 2 and soybean mosaic virus-N. *J Gen Virol*. 1989; 70: 2775–2783. <https://doi.org/10.1099/0022-1317-70-10-2775> PMID: 2794980
67. Adams MJ, Antoniw JF, Fauquet CM. Molecular criteria for genus and species discrimination within the family Potyviridae. *Arch Virol*. 2005; 150: 459–479. <https://doi.org/10.1007/s00705-004-0440-6> PMID: 15592889
68. Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li B-C, Herrmann R. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res*. 1996; 24: 4420–4449. <https://doi.org/10.1093/nar/24.22.4420> PMID: 8948633
69. Sonnhammer ELL, Von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Ismb*. 1998. pp. 175–182. PMID: 9783223
70. Acharya V, Chakraborty HJ, Rout AK, Balabantaray S, Behera BK, Das BK. Structural characterization of open reading frame-encoded functional genes from tilapia lake virus (TiLV). *Mol Biotechnol*. 2019; 61: 945–957. <https://doi.org/10.1007/s12033-019-00217-y> PMID: 31664705
71. Ginalski K. Comparative modeling for protein structure prediction. *Curr Opin Struct Biol*. 2006; 16: 172–177. <https://doi.org/10.1016/j.sbi.2006.02.003> PMID: 16510277
72. Rout AK, Paramanik S, Dehury B, Acharya V, Swain HS, Pradhan SK, et al. Elucidating the molecular interaction of Zebrafish (*Danio rerio*) peptidoglycan recognition protein 2 with diaminopimelic acid and lysine type peptidoglycans using in silico approaches. *J Biomol Struct Dyn*. 2020; 38: 3687–3699. <https://doi.org/10.1080/07391102.2019.1666742> PMID: 31517586
73. Rout AK, Acharya V, Maharana D, Dehury B, Udgata SR, Jena R, et al. Insights into structure and dynamics of extracellular domain of Toll-like receptor 5 in *Cirrhinus mrigala* (*mrigala*): A molecular dynamics simulation approach. *PLoS One*. 2021; 16: e0245358. <https://doi.org/10.1371/journal.pone.0245358> PMID: 33444377
74. Chakraborty HJ, Rout AK, Behera BK, Parhi J, Parida PK, Das BK. Insights into the aquaporin 4 of zebrafish (*Danio rerio*) through evolutionary analysis, molecular modeling and structural dynamics. *Gene Reports*. 2018; 11: 101–109.
75. Rout AK, Udgata SR, Dehury B, Pradhan SP, Swain HS, Behera BK, et al. Structural bioinformatics insights into the CARD-CARD interaction mediated by the mitochondrial antiviral-signaling protein of black carp. *J Cell Biochem*. 2019; 120: 12534–12543. <https://doi.org/10.1002/jcb.28519> PMID: 30912187
76. Dehury B, Behera SK, Mahapatra N. Structural dynamics of Casein Kinase I (CKI) from malarial parasite *Plasmodium falciparum* (Isolate 3D7): Insights from theoretical modelling and molecular simulations. *J Mol Graph Model*. 2017; 71: 154–166. <https://doi.org/10.1016/j.jmgl.2016.11.012> PMID: 27923179

77. Nayak C, Chandra I, Singh SK. An in silico pharmacological approach toward the discovery of potent inhibitors to combat drug resistance HIV-1 protease variants. *J Cell Biochem.* 2019; 120: 9063–9081. <https://doi.org/10.1002/jcb.28181> PMID: 30506751
78. Rout AK, Dehury B, Maharana J, Nayak C, Baisvar VS, Behera BK, et al. Deep insights into the mode of ATP-binding mechanism in Zebrafish cyclin-dependent protein kinase-like 1 (zCDKL1): A molecular dynamics approach. *J Mol Graph Model.* 2018; 81: 175–183. <https://doi.org/10.1016/j.jmgl.2018.02.002> PMID: 29574323
79. Rout AK, Mishra J, Dehury B, Maharana J, Acharya V, Karna SK, et al. Structural bioinformatics insights into ATP binding mechanism in zebrafish (*Danio rerio*) cyclin-dependent kinase-like 5 (zCDKL5) protein. *J Cell Biochem.* 2019; 120: 9437–9447. <https://doi.org/10.1002/jcb.28219> PMID: 30569538
80. Behera BK, Parhi J, Dehury B, Rout AK, Khatei A, Devi AL, et al. Molecular characterization and structural dynamics of Aquaporin1 from walking catfish in lipid bilayers. *Int J Biol Macromol.* 2022; 196: 86–97. <https://doi.org/10.1016/j.ijbiomac.2021.12.014> PMID: 34914911
81. Roy P, Rout AK, Maharana J, Sahoo DR, Panda SP, Pal A, et al. Molecular characterization, constitutive expression and GTP binding mechanism of *Cirrhinus mrigala* (Hamilton, 1822) Myxovirus resistance (Mx) protein. *Int J Biol Macromol.* 2019; 136: 1258–1272. <https://doi.org/10.1016/j.ijbiomac.2019.06.161> PMID: 31242450
82. Yesylevskyy SO, Kharkyanen VN, Demchenko AP. The change of protein intradomain mobility on ligand binding: is it a commonly observed phenomenon? *Biophys J.* 2006; 91: 3002–3013. <https://doi.org/10.1529/biophysj.106.087866> PMID: 16877502