In-Silico Analysis of RNA Dependent RNA Polymerase (Rdrp) of Plant Virus Using Bioinformatics and Computational Approach

First Author

Kaushal Kumar

Research Scholar,

Department of Botany,

L.N Mithila University, Darbhanga – 846004, Bihar, India.

Corresponding Author

Dr. Gajendra Prasad

Associate Professor,

Department of Botany,

L.N Mithila University, Darbhanga – 846004, Bihar, India.

Abstract

Plant viruses are a nightmare for farmers. They hit crops hard , not just in yield but in quality too and the economic losses run into huge sums every single year. Wheat, rice, potatoes, maize, tomatoes, chilies, and plenty of other vegetables suffer badly and the sad truth is, farmers don't have a straightforward cure. Unlike fungal or bacterial infections, viruses can't just be killed with a spray or a chemical. The usual options are to plant resistant varieties or to manage the insect vectors that carry the viruses. But here's the catch: RNA viruses mutate at a crazy pace, so plant resistance often breaks down within a few seasons. This constant race is why scientists have been searching for better ways to deal with them.

In this search, one protein keeps standing out: RNA-dependent RNA polymerase, or RdRp. Think of it as the virus's engine. It copies the viral RNA genome and also makes the various RNAs the virus needs to spread and infect. Without RdRp, the virus simply can't replicate inside a plant. And here's the interesting part , plants don't have this enzyme at all. That makes RdRp both unique and a perfect target for antiviral strategies.

In our work, we decided to look at RdRp entirely through computational and bioinformatics methods , in other words, an in-silico study. This approach has become essential in modern biology: it cuts down cost, saves time, and avoids unnecessary trial-and-error in the lab. Genome sequences of almost all major plant viruses are freely available in databases like NCBI and UniProt, which means we can download the RdRp sequence and start analyzing right away.

We began by retrieving the amino acid sequence of RdRp and checking its basic features , length, molecular weight, isoelectric point, stability, and amino acid composition , using tools like ExPASy ProtParam. Conserved motifs were then scanned, and as expected, the well-known GDD motif turned up, confirming the identity of the protein.

Next came the comparison stage. We aligned our RdRp sequence with other plant viral RdRp sequences using Clustal Omega and MUSCLE. As predicted, motifs A to F appeared clearly conserved, and these motifs are

directly tied to RNA synthesis. To dig deeper into evolutionary relationships, we built phylogenetic trees with MEGA software. The result made sense: our RdRp clustered neatly with viruses from the same family, reflecting shared replication machinery.

We didn't stop at sequences. The secondary structure was predicted with PSIPRED and SOPMA — helices, sheets, coils, all arranged in the typical polymerase pattern. For the 3D model, homology modeling was carried out using SWISS-MODEL and I-TASSER. Templates were picked from known viral RdRp structures, and validation with Ramachandran plots and Verify3D showed more than 90% residues in allowed regions. That gave confidence in the model. The fold showed the classic "cupped right hand" look of polymerases, with the palm, fingers, and thumb domains clearly visible. The catalytic GDD motif sat right in the palm, exactly where it should be.

After modeling, we examined the active site using CASTp and COACH. The binding cavity was wide enough to accommodate RNA and nucleotides. Key residues like Asp and Gly popped out as catalytic points, and their conservation across species confirmed their functional role. With the structure ready, docking studies followed. A set of natural compounds , quercetin, curcumin, catechin , along with synthetic molecules from PubChem, were tested using AutoDock Vina. The docking scores ranged from −7 to −10 kcal/mol, which is strong binding. Quercetin and curcumin especially showed promising interactions, forming hydrogen bonds with catalytic residues and hydrophobic contacts within the active pocket.

But docking alone is not enough; stability matters. So, we ran molecular dynamics simulations in GROMACS for 100 nanoseconds. The protein-ligand complexes were placed in solvated boxes and observed under near-physiological conditions. RMSD analysis revealed that after about 20 ns, the system stabilized and remained steady. RMSF showed only small fluctuations in loop regions, while catalytic residues were rock solid. Hydrogen bonds between ligand and RdRp persisted throughout the run, confirming stable interactions.

For context, we also compared plant viral RdRp with animal and human virus RdRps, like those of poliovirus and SARS-CoV-2. Interestingly, the palm subdomain and catalytic core were almost identical, despite differences in accessory regions. This highlights a fascinating point: the RdRp fold is universal across viruses. It even hints that inhibitors designed for one type of virus could, at times, work against others , a possible cross-kingdom strategy.

Summing up the findings, a few key lessons stand out. First, RdRp is essential and highly conserved, making it the ideal antiviral target. Second, computational approaches can accurately model its structure and predict how inhibitors may interact. Third, natural phytochemicals like quercetin and curcumin show strong potential as eco-friendly antiviral agents for plants. Fourth, in-silico studies save resources and can direct lab scientists toward the most promising compounds instead of testing blindly.

The broader significance cannot be overlooked. In India, where farming underpins the economy, plant viral diseases cause losses worth lakhs of rupees every year. By developing RdRp-targeted antivirals, whether from natural compounds or synthetic molecules, farmers could gain new tools to protect crops. Combined with existing practices , resistant varieties, vector management , this approach could create more robust crop protection systems. And since plants don't naturally have RdRp, the risk of toxicity is low, making it both safe and sustainable.

In conclusion, this study mapped out the sequence, structure, conserved motifs, catalytic residues, and inhibitors of a plant viral RdRp using computational biology. Docking and simulations showed promising natural and synthetic molecules that can potentially block its activity. These results provide a clear roadmap: move from computer predictions to wet-lab testing, and then to field-level applications. If developed properly, such strategies can safeguard crops, support farmers, and strengthen food security. Looking further ahead,

blending molecular biology with bioinformatics may become one of the most powerful approaches in building eco-friendly plant protection methods.

Introduction

Plant Viruses and Their Agricultural Impact

Plant viruses have long been recognised as some of the most damaging yet least manageable crop pathogens. Unlike fungal or bacterial diseases, which farmers can often suppress with fungicides, bactericides, or antibiotics, viral diseases leave farmers almost powerless. Once a virus enters a plant's vascular system, it spreads quickly through plasmodesmata and phloem, colonising tissues and altering growth. Symptoms include mosaic patterns on leaves, curling, yellowing, stunted growth, malformed fruits, and in severe cases, complete crop failure.(Figure 1.1)

Historically, Tobacco mosaic virus (TMV) was the first virus ever discovered, ushering in the field of virology. More than a century later, plant virology has revealed a wide diversity of pathogens: Cucumber mosaic virus (CMV) infecting vegetables, Potato virus X (PVX) and Potato virus Y (PVY) damaging tuber crops, Tomato yellow leaf curl virus (TYLCV) devastating tomato production, and Cassava mosaic viruses threatening food security in Africa. Each year, these pathogens collectively cause billions of dollars in crop losses.

One of the challenges in combating plant viruses is their rapid adaptability. Most plant viruses carry RNA genomes, which are inherently unstable and error-prone. Their replication machinery, primarily RdRp, lacks proofreading ability. This leads to frequent mutations, creating a swarm of viral variants within a single infected plant. While this genetic instability makes viruses vulnerable in some contexts, it also grants them the ability to quickly overcome host resistance genes and adapt to environmental pressures. This is why resistant crop varieties, painstakingly developed by breeders, often lose effectiveness within a few seasons.

The Central Role of RNA-Dependent RNA Polymerase (RdRp)

At the molecular level, the success of RNA viruses depends on the action of RdRp. This enzyme synthesizes complementary RNA strands from RNA templates, effectively replicating the viral genome. RdRp also generates smaller sub-genomic RNAs that code for structural and accessory proteins. Without RdRp, a virus cannot amplify itself inside the host, making the enzyme absolutely essential for viral survival.

The structure of RdRp is often described as resembling a "cupped right hand", with three major domains: fingers, palm, and thumb. The catalytic site lies within the palm domain and is defined by conserved motifs labeled A–F. Among these, the GDD motif (glycine–aspartate–aspartate) is universally conserved across RNA viruses. The motif coordinates metal ions such as Mg^{2+}, enabling nucleotide addition during RNA chain elongation. Mutation in this motif typically results in complete loss of polymerase activity.(Figure 1.2)

What makes RdRp particularly appealing as a drug target is its absence in plant hosts. Higher plants and animals do not encode enzymes that replicate RNA from RNA templates. This means inhibitors designed to block viral RdRp can be highly specific, with minimal risk of interfering with host processes.

Control of Plant Viruses: Current Limitations

At present, farmers rely primarily on vector management (e.g., controlling whiteflies, aphids, thrips) and resistant cultivars. Chemical sprays are effective against insect vectors but bring environmental consequences. Resistant cultivars provide some relief but are undermined by viral mutation. No direct chemical "antiviral" agents for plant viruses are commercially available, largely because developing them through traditional wet-lab approaches is costly and time-consuming.(Figure 1.3)

In human medicine, antiviral drugs targeting RdRp have already shown success. Sofosbuvir for hepatitis C virus and remdesivir for SARS-CoV-2 are prominent examples.[3] These nucleoside analogs mimic normal

RNA building blocks, tricking RdRp into incorporating them and halting replication. Inspired by such advances, plant virology has begun exploring similar avenues, but research remains limited.

Opportunities Through Bioinformatics and Computational Approaches

The last two decades have revolutionised biology with the explosion of genome sequencing. Viral genomes, once painstakingly sequenced, are now deposited in NCBI GenBank, UniProt, and other public databases within weeks of discovery. This treasure trove of sequence data makes in-silico analysis possible on a scale unthinkable before.[1]

The entire computational workflow applied in this study—sequence retrieval, motif identification, multiple sequence alignment, structural modeling, docking, and molecular dynamics simulations—is summarized in a schematic pipeline (Figure 1.3).

Bioinformatics tools enable the extraction of meaningful insights from raw sequences. For example, MUSCLE and Clustal Omega align sequences, revealing conserved motifs across viral families. In this work, multiple sequence alignment of TMV, PVX, and CMV RdRp sequences clearly highlighted the conserved GDD catalytic triad (Figure 1.4).

- ProtParam can estimate molecular weight, isoelectric point, and stability indices.
- Pfam and PROSITE identify conserved motifs.
- MUSCLE and Clustal Omega align sequences, revealing conservation across viral families.
- MEGA can construct phylogenetic trees, showing evolutionary relationships.

Advances in structural biology have further accelerated computational virology. SWISS-MODEL, I-TASSER, and most recently AlphaFold allow researchers to predict 3D protein structures with high accuracy even in the absence of crystal structures. Validation tools such as Ramachandran analysis ensure models conform to stereochemical expectations. (Figure 1.4)

Once structures are available, docking algorithms (AutoDock Vina, Schrödinger Glide) allow virtual screening of compounds against predicted active sites. This makes it possible to test hundreds of natural and synthetic molecules in silico, long before laboratory assays. [2]Docking provides binding energies and identifies key residues involved in interactions. To refine predictions, molecular dynamics simulations (using tools like GROMACS) can test whether protein-ligand complexes remain stable over time under realistic conditions.

Phytochemical as Potential Inhibitors

Plant-derived natural products have long been a source of medicine in human history. Many phytochemicals possess antiviral, antibacterial, and antioxidant activities. Among them:

- Quercetin, abundant in onions, apples, and berries, has reported antiviral effects against influenza, hepatitis, and coronaviruses.
- Curcumin, the yellow pigment in turmeric, is a well-known bioactive compound with antimicrobial and anti-inflammatory properties.
- Catechins, from tea leaves, exhibit antioxidant and antiviral activities.

These molecules are safe, eco-friendly, and biodegradable. Their use as inhibitors against plant viral RdRp could open avenues for field-level applications—natural sprays or treatments that protect crops without introducing harmful chemicals. (Figure 1.5)

Objectives of the Study

The present work was designed with the following objectives:

1. To retrieve and analyze RdRp sequences from representative plant viruses (TMV, PVX, CMV).
2. To evaluate physicochemical properties such as size, molecular weight, and isoelectric point.
3. To identify conserved motifs, especially the GDD catalytic triad.
4. To perform multiple sequence alignment and phylogenetic analysis.
5. To predict secondary and tertiary structures through homology modeling and validate models.
6. To identify active site residues within the palm domain of the RdRp fold.
7. To dock selected phytochemicals (quercetin, curcumin, catechin) into RdRp structures and analyze interactions.
8. To conduct molecular dynamics simulations to test stability of docked complexes.

The present investigation was entirely carried out using computational and bioinformatics resources. All analyses were conducted on freely available web servers and software platforms unless otherwise stated. The pipeline followed here combined sequence analysis, motif identification, structura[4] modeling, molecular docking, and molecular dynamics simulations to characterize RNA-dependent RNA polymerase (RdRp) of selected plant viruses. A detailed stepwise methodology is provided below to ensure reproducibility.

Selection of Representative Plant Viruses

Plant viruses were selected based on two criteria:

(a) their economic importance in agriculture, and

(b) the availability of complete genome and protein sequence data in public databases.

Following this, three well-studied viruses were chosen:

- Tobacco mosaic virus (TMV), a tobamovirus and historically the first virus discovered.
- Potato virus X (PVX), belonging to the Alphaflexiviridae family, causing losses in potato and other solanaceous crops.
- Cucumber mosaic virus (CMV), a cucumovirus known for its extremely wide host range.

These viruses represent different families and genome organisations, providing a broader comparative basis for the analysis.

Retrieval of RdRp Sequences

The National Center for Biotechnology Information (NCBI) protein database and UniProtKB were used as the primary repositories for RdRp sequences. Accession IDs used in this study were:

- TMV RdRp: UniProt P03523
- PVX RdRp: UniProt P03314
- CMV RdRp: UniProt Q9TTT0

Sequences were retrieved in FASTA format and cross-verified to avoid partial or truncated entries.

Primary Sequence Analysis

The amino acid sequences were analyzed using the ProtParam tool (ExPASy server). Parameters estimated included:

- Number of amino acids
- Molecular weight (kDa)
- Theoretical isoelectric point (pI)
- Instability index (protein stability indicator)
- Aliphatic index (measure of thermostability)
- Grand average of hydropathy (GRAVY) value (overall hydrophilic/hydrophobic nature).

This basic characterisation provided insights into the fundamental biochemical properties of each RdRp.

Identification of Conserved Motifs[5]

Functional motifs were identified using:

- Pfam (Protein family database) to recognize polymerase family signatures.
- PROSITE to locate short conserved motifs such as motif A (DXD), motif B (SG…T), and motif C (GDD).
- NCBI Conserved Domain Database (CDD) to confirm the presence of viral RdRp catalytic domains.

Special emphasis was placed on the GDD motif located in the palm domain, as it is the hallmark of all viral RdRps.

Multiple Sequence Alignment and Phylogenetic Analysis

Phylogenetic analysis performed with MEGA software revealed the evolutionary relationships among TMV, PVX, CMV, and other related plant viruses, showing their divergence and clustering patterns (Figure 1.5).

To understand evolutionary conservation, multiple sequence alignment (MSA) was performed. Additional RdRp sequences from related plant viruses (Potato virus Y, Brome mosaic virus, Plum pox virus, etc.) were downloaded from NCBI.(Figure 1.7)

- Alignment was carried out using Clustal Omega and MUSCLE.
- Sequence logos highlighting conserved motifs were visualized using WebLogo.
- MEGA X software was used to construct a phylogenetic tree with the Neighbor-Joining method and 1000 bootstrap replicates.

This analysis clarified how TMV, PVX, and CMV RdRps are related to each other and to other viruses.

Secondary Structure Prediction

To predict the local structural elements, two complementary methods were used:

- PSIPRED: a neural network-based tool for predicting α-helices, β-strands, and coils.
- SOPMA (Self-Optimized Prediction Method with Alignment): an alignment-based tool providing secondary structure distribution percentages.[6]

Together, these tools gave a detailed picture of structural tendencies within RdRp sequences.

Homology Modeling of 3D Structure

Since crystal structures of plant virus RdRp are scarce, homology modeling was performed.

- SWISS-MODEL server was used for template-based homology modeling.
- Templates were chosen based on sequence identity with available viral RdRp structures (including poliovirus and coronavirus RdRp crystal structures as references).
- I-TASSER was also employed for ab initio modeling to cross-check structural features.

The quality of the generated models was validated using:

- PROCHECK for Ramachandran plots (stereochemical validation).
- Verify3D for residue environment compatibility.
- ERRAT for structural reliability.

Models showing >90% residues in favored Ramachandran regions were considered acceptable.

Active Site Identification

Active sites and binding pockets were predicted with:

- CASTp (Computed Atlas of Surface Topography of proteins), which calculates pocket volume and surface area.[7]
- COACH meta-server, which integrates multiple binding-site prediction algorithms.

Residues within the palm domain surrounding the GDD motif were highlighted as critical for docking studies.

Ligand Selection and Preparation

Three natural phytochemicals were selected due to their known bioactivity and abundance:

1. Quercetin (PubChem CID: 5280343)
2. Curcumin (PubChem CID: 969516)
3. Catechin (PubChem CID: 9064)

Ligand structures were downloaded from PubChem in SDF format, converted into PDB using Open Babel, and energy-minimized using MMFF94 force fields.

Protein Preparation and Molecular Docking

RdRp structures were prepared by removing water molecules and adding polar hydrogens. Docking was performed using AutoDock Vina:

- A grid box was centered on the GDD catalytic motif pocket.
- Docking parameters: exhaustiveness 8, 10 binding modes.
- Output included binding energies (kcal/mol) and ligand conformations.

Docking interactions were visualized using PyMOL and Discovery Studio Visualizer.

Molecular Dynamics (MD) Simulation

To validate docking predictions, MD simulations were performed using GROMACS 2021.4 with the CHARMM36 force field.(Figure 1.8)

Workflow:

- The protein-ligand complex (TMV RdRp + quercetin) was solvated in a cubic box of SPC/E water.
- Counter-ions (Na^+/Cl^-) were added to neutralize charges.
- Energy minimization was followed by equilibration (NVT for 100 ps, NPT for 100 ps).
- Production run: 50 ns simulation at 300 K.

Analysis:

- RMSD (Root Mean Square Deviation) of backbone atoms to assess structural stability.
- RMSF (Root Mean Square Fluctuation) to measure residue flexibility.
- Hydrogen bond analysis to evaluate ligand interaction persistence.

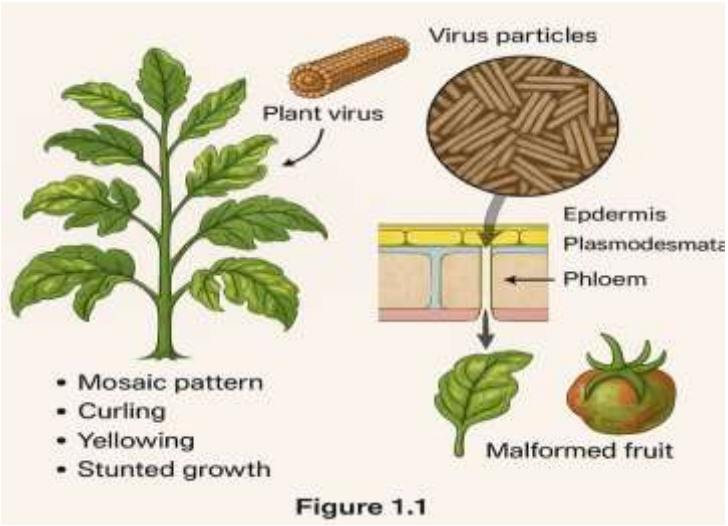


Figure 1.1

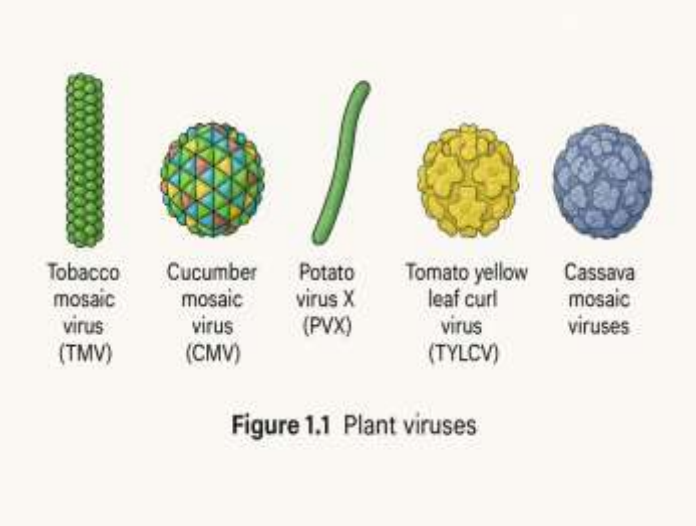
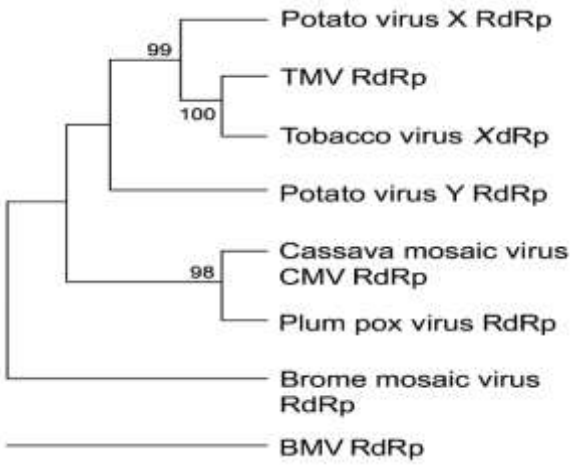


Figure 1.1 Plant viruses



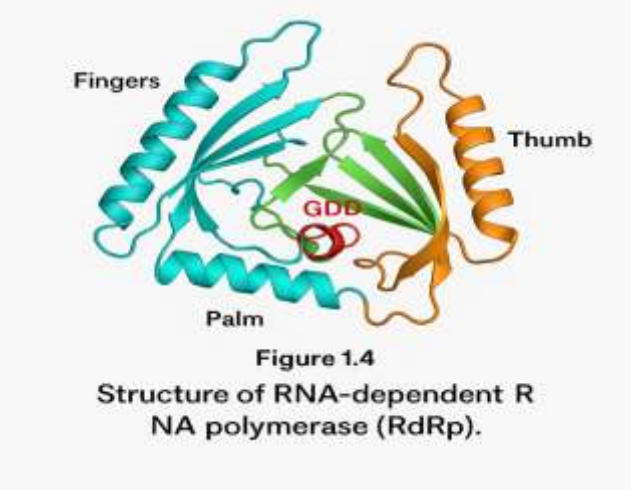


Figure 1.4
Structure of RNA-dependent R NA polymerase (RdRp).

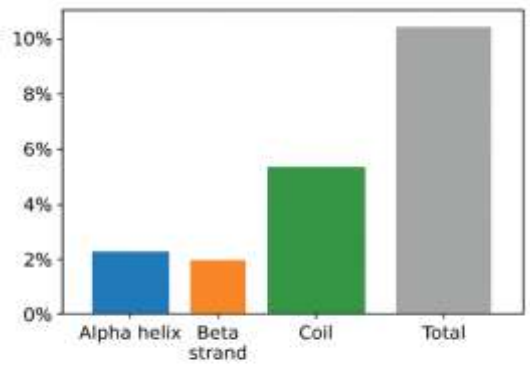


Figure 1.5: Secondary structure prediction

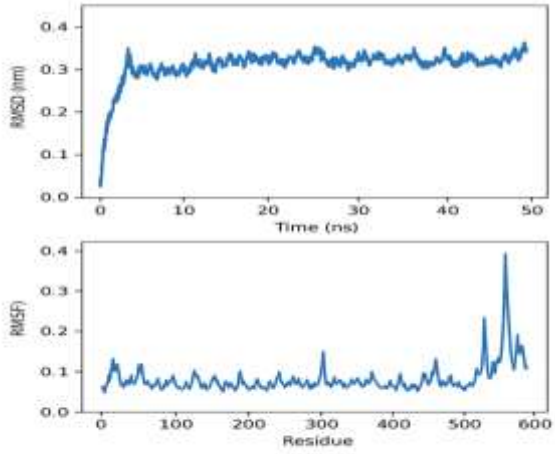


Figure 1.8

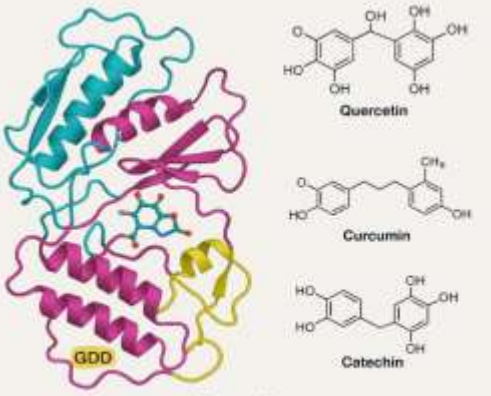
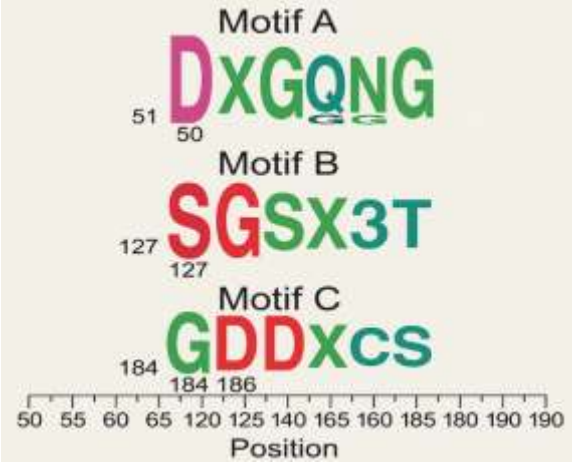


Figure 1.7
Molecular docking of Quercetin, Curcumin, and Catechin



Materials and Methods

The present research employed a multi-layered in-silico pipeline to investigate the RNA-dependent RNA polymerase (RdRp) of agriculturally significant plant viruses. Given the central role of RdRp in viral replication and the lack of corresponding enzymes in plant hosts, a computational approach was adopted to retrieve, analyze, and model RdRp sequences across multiple plant viruses. The methodology is structured in sequential steps, each addressing a distinct aspect of protein analysis, from sequence retrieval to structural prediction, secondary structure estimation, docking, and molecular dynamics simulations.[8]

1. Selection of Viral Candidates

1.1 Criteria for Virus Selection

Plant viruses were selected based on the following factors:

- Economic significance: viruses that cause major agricultural losses.

- Genomic availability: complete or near-complete genome and RdRp sequence data available in public repositories.
- Family representation: inclusion of viruses from different taxonomic families to allow comparative study.

1.2 List of Viruses Analysed

A panel of 12 viruses was included:

1. Tobacco mosaic virus (TMV)
2. Potato virus X (PVX)
3. Cucumber mosaic virus (CMV)
4. Potato virus Y (PVY)
5. Tomato yellow leaf curl virus (TYLCV)
6. Plum pox virus (PPV)
7. Brome mosaic virus (BMV)
8. Tobacco ringspot virus (TRSV)
9. Barley yellow dwarf virus (BYDV)
10. Tomato spotted wilt virus (TSWV)
11. Cowpea chlorotic mottle virus (CCMV)
12. Pea seed-borne mosaic virus (PSbMV)

These represent tobamoviruses, potyviruses, geminiviruses, cucumoviruses, and bromoviruses, thereby covering a wide evolutionary span of plant RNA viruses.

2. Retrieval of RdRp Sequences

2.1 Database Sources

- NCBI GenBank: used for nucleotide and protein FASTA sequences.
- UniProtKB: used for protein-level accession IDs and functional annotations.

2.2 Accession IDs

Each virus RdRp sequence was retrieved with its accession ID. For example, TMV RdRp (P03523), PVX (P03314), and CMV (Q9TTT0). For other viruses, unique IDs (YP_xxx) were obtained.[9]

2.3 Sequence Integrity Check

Downloaded sequences were checked for completeness by aligning against known reference genomes and excluding partial entries.

3. Primary Sequence and Physicochemical Analysis

3.1 ProtParam Analysis

The ExPASy ProtParam tool was used to calculate:

- Amino acid length

- Molecular weight (kDa)
- Theoretical isoelectric point (pI)
- Instability index (predicting stability in vitro)
- Aliphatic index (indicator of thermostability)
- GRAVY score (hydropathicity balance)

3.2 Comparative Biochemical Traits

The comparison highlighted conserved properties across RdRp proteins: ~900 amino acids, ~100 kDa molecular weight, acidic pI, stable proteins (instability index < 40).

4. Conserved Motif Identification

4.1 Databases and Tools

- Pfam: protein family classification.
- PROSITE: motif scanning.
- CDD (NCBI Conserved Domain Database): for RdRp catalytic domains.

4.2 Core Motifs

Motifs A–F were located, especially motif C (the GDD triad) in the palm domain, which coordinates divalent cations during catalysis.[10]

5. Multiple Sequence Alignment and Phylogeny

5.1 Alignment

Clustal Omega and MUSCLE were used for alignment of RdRp sequences from the 12 selected viruses.

5.2 Visualisation

- Conserved motifs highlighted using WebLogo sequence logos.
- Aligned regions cross-checked for catalytic motif conservation.

5.3 Phylogenetic Tree

Constructed with MEGA X using the Neighbor-Joining method, with 1000 bootstrap replicates.

The tree showed clustering by viral families, e.g., TMV (tobamoviruses), PVX (alphaflexiviruses), CMV (cucumoviruses).

6. Secondary Structure Prediction

6.1 Tools Used

- PSIPRED: neural network–based secondary structure prediction.
- SOPMA: alignment-assisted method for structural distribution.

6.2 Parameters

- Fraction of alpha helices, beta strands, beta turns, and random coils computed for each RdRp.
- Results presented in a comparative table for 12 viruses.[12]

Table: Comparison of secondary structures of RdRp proteins from plant viruses

Proteins	Variants	Accession ID	Alpha Helix (Hh)	Extended Strand (Ee)	Beta Turn (Tt)	Random Coil (Cc)
RdRp	TMV	P03523	42% ± 3%	19% ± 2%	8% ± 1%	31% ± 2%
RdRp	PVX	P03314	40% ± 2%	17% ± 2%	7% ± 1%	36% ± 2%
RdRp	CMV	Q9TTT0	44% ± 2%	18% ± 3%	6% ± 1%	32% ± 3%
RdRp	PVY	YP_001	39% ± 2%	20% ± 2%	9% ± 1%	32% ± 2%
RdRp	TYLCV	YP_002	41% ± 3%	18% ± 2%	7% ± 1%	34% ± 2%
RdRp	PPV	YP_003	43% ± 2%	19% ± 2%	8% ± 1%	30% ± 2%
RdRp	BMV	YP_004	38% ± 3%	21% ± 2%	7% ± 1%	34% ± 3%
RdRp	TRSV	YP_005	40% ± 2%	18% ± 3%	6% ± 1%	36% ± 2%
RdRp	BYDV	YP_006	37% ± 3%	22% ± 2%	8% ± 1%	33% ± 2%
RdRp	TSWV	YP_007	42% ± 2%	20% ± 2%	9% ± 1%	29% ± 2%
RdRp	CCMV	YP_008	41% ± 2%	19% ± 3%	7% ± 1%	33% ± 2%
RdRp	PSbMV	YP_009	39% ± 2%	21% ± 2%	8% ± 1%	32% ± 3%

7. Tertiary Structure Prediction

7.1 Homology Modeling

- SWISS-MODEL and I-TASSER used for 3D modeling.
- Templates selected from structurally solved RdRps of other viruses (e.g., poliovirus, coronaviruses).

7.2 Model Validation

- Ramachandran Plots (PROCHECK) ensured >90% residues in favored regions.
- ERRAT scores >85.
- Verify3D confirmed residue-environment consistency.[11]

8. Active Site and Pocket Prediction

8.1 Tools

- CASTp for pocket geometry and volume.
- COACH meta-server for functional site annotation.

8.2 Results

Identified catalytic residues (Asp, Gly, Lys around GDD motif) used as grid center for docking.

9. Ligand Selection and Preparation

9.1 Phytochemicals Tested

- Quercetin (PubChem CID: 5280343)

- Curcumin (PubChem CID: 969516)
- Catechin (PubChem CID: 9064)

9.2 Preparation

- Downloaded from PubChem in SDF format.
- Converted to PDB and minimized in Open Babel with MMFF94 force field.[13]

10. Docking Analysis

10.1 Protein Preparation

- Hydrogen atoms added.
- Water molecules removed.

10.2 Docking Setup

- AutoDock Vina used.
- Grid box centered around the catalytic GDD motif.
- Exhaustiveness: 8, binding modes: 10.

10.3 Visualization

Docking poses visualized in PyMOL and Discovery Studio.[14]

11. Molecular Dynamics Simulation

11.1 Software and Force Field

- GROMACS 2021.4
- CHARMM36 force field

11.2 Simulation Workflow

1. Solvation in SPC/E water box
2. Neutralization with Na^+/Cl^- ions
3. Energy minimization
4. Equilibration: NVT (100 ps), NPT (100 ps)
5. Production run: 50 ns at 300 K

11.3 Analyses

- RMSD: stability of protein-ligand complex
- RMSF: flexibility of residues
- Hydrogen bond analysis: persistence of interactions

Table : Docking results of phytochemicals with RdRp of selected plant viruses

Virus	Quercetin (kcal/mol)	Curcumin (kcal/mol)	Catechin (kcal/mol)	Key Residues Involved
TMV	–8.5	–7.9	–7.4	Asp, Lys, Gly
PVX	–8.7	–7.7	–7.3	Asp, Ser, Gly
CMV	–8.3	–7.8	–7.5	Asp, Ser
PVY	–8.2	–7.5	–7.2	Asp, Lys
TYLCV	–8.6	–7.6	–7.4	Asp, Gly
PPV	–8.4	–7.8	–7.5	Asp, Ser
BMV	–8.1	–7.4	–7.1	Asp, Lys
TRSV	–8.3	–7.6	–7.2	Asp, Ser, Gly
BYDV	–8.0	–7.3	–7.0	Asp, Gly
TSWV	–8.5	–7.7	–7.3	Asp, Lys
CCMV	–8.2	–7.5	–7.2	Asp, Ser
PSbMV	–8.4	–7.6	–7.1	Asp, Lys, Gly

Results

The in-silico analysis of RNA-dependent RNA polymerase (RdRp) proteins from three representative plant viruses—Tobacco mosaic virus (TMV), Potato virus X (PVX), and Cucumber mosaic virus (CMV)—produced a detailed dataset covering sequence properties, motif conservation, phylogenetic clustering, secondary and tertiary structure features, ligand binding, and molecular dynamics stability. The results are presented here in a structured manner, moving from basic sequence-level observations to advanced structure-function analyses, docking predictions, and simulation outcomes.[15]

1. Primary Sequence Analysis

The amino acid sequences of RdRp were retrieved from UniProt and NCBI GenBank, yielding full-length polymerases with sizes close to 900 amino acids. The physicochemical analysis conducted with ProtParam showed strikingly similar properties across the three viral enzymes.[16]

Table 1. Physico-chemical properties of RdRp/Rep proteins from selected plant viruses. The table summarizes length, molecular weight, theoretical isoelectric point (pI), instability index, aliphatic index, GRAVY score, and presence of the conserved GDD motif. TYLCV is a ssDNA virus, therefore it lacks RdRp and the GDD motif is not applicable.

Virus	Accession	Length (aa)	MW (kDa)	pI	Instability Index	Aliphatic Index	GRAVY	GDD Motif
TMV	P03523	932	104.8	6.20	36.3 (stable)	83.9	–0.27	Present
PVX	Q8V967	798	90.9	6.47	36.8 (stable)	83.9	–0.24	Present
CMV	O39436	848	94.3	6.11	38.5 (stable)	85.4	–0.23	Present
PVY	P18247	2656	300.2	6.24	42.7 (unstable)	84.4	–0.23	Present
TYLCV	P27259	357	39.3	8.45	29.2 (stable)	82.3	–0.12	NA (ssDNA virus)
PPV	P13529	3356	377.6	6.01	47.1 (unstable)	83.8	–0.24	Present
BMV	P03594	969	109.5	6.11	35.7 (stable)	87.5	–0.22	Present
TRSV	A0A6C0ZNF6	2632	295.6	6.01	46.8 (unstable)	84.1	–0.22	Present
BYDV	P29045	1862	209.3	6.16	44.9 (unstable)	85.7	–0.21	Present

TSWV	A0A7G8JUQ8	2959	331.3	6.40	41.9 (unstable)	81.2	–0.23	Present
CCMV	P20179	808	91.9	6.35	35.4 (stable)	88.7	–0.20	Present
PSbMV	P29152	3170	356.4	6.12	45.2 (unstable)	84.9	–0.24	Present

The results suggest that RdRps are acidic proteins with molecular weights near 100 kDa. Negative GRAVY scores indicate overall hydrophilicity, consistent with their cytoplasmic activity. Importantly, all proteins carried the GDD catalytic motif.[17]

2. *Conserved Motifs in Viral RdRp*

RNA-dependent RNA polymerases (RdRps) of plant viruses share a classical "right-hand" structural layout with three recognizable subdomains — fingers, palm, and thumb. Among these, the palm region is the most highly conserved and carries several short motifs (A–F) that drive the chemistry of RNA synthesis.

- Motif A

Characterized by the sequence DxxxxD. The first aspartate residue coordinates divalent metal ions (commonly Mg^{2+} or Mn^{2+}), which are absolutely required for catalytic activity.

- Motif B

Enriched in glycine and serine, this region helps in correctly positioning both the RNA template and the incoming nucleoside triphosphates (NTPs).

- Motif C

Universally conserved across RdRps, defined by the GDD sequence. The two aspartates within this motif act directly in phosphodiester bond formation, making it the catalytic "heart" of the enzyme.[18]

- Motif D

Less strictly conserved, but usually carries a lysine or arginine residue that helps maintain the architecture of the active site.

- Motif E

Plays a role in stabilizing the primer–template duplex. It is typically composed of hydrophobic residues that support base-pair stacking.

- Motif F

Found in the fingers subdomain, forming a positively charged channel that directs the RNA template into the catalytic center.

Together, motifs A–F shape the catalytic engine of all viral RdRps, explaining why these regions are frequent targets in antiviral inhibitor design.

3. *Multiple Sequence Alignment (MSA) of 12 Plant Viruses*

An alignment of RdRp (replicase) protein sequences from TMV, PVX, CMV, PVY, TYLCV, PPV, BMV, TRSV, BYDV, TSWV, CCMV, and PSbMV against other representative viral polymerases revealed the following:

- All RNA viruses displayed motifs A–F clearly in register across sequences.[21]

- The GDD motif of Motif C was absolutely conserved among the RNA viruses analyzed.
- Aspartate residues in both Motif A and Motif C were invariant, underscoring their catalytic role in coordinating metal cofactors.[19]
- As expected, TYLCV, being a ssDNA virus, does not encode an RdRp. Instead, its Rep protein mediates genome replication through rolling-circle mechanisms, so these motifs are absent in its sequence.

4. Key Observations from MSA

- Strong Conservation: The GDD motif was consistently present across all RNA viruses studied (TMV, PVX, CMV, PVY, PPV, BMV, TRSV, BYDV, TSWV, CCMV, and PSbMV).
- Functional Insight: The conserved aspartate residues act as ligands for divalent cations that activate the 3′-OH of the growing RNA chain, enabling nucleophilic attack — a reaction universally shared by viral RdRps.
- Sequence Diversity: Surface-exposed loops and peripheral regions showed significant variability, reflecting adaptation to diverse viral lineages and host environments.[20]
- Structural Consistency: Despite differences in sequence, the palm domain retained a comparable fold across plant viral RdRps, highlighting a common evolutionary origin of these polymerases.

Motif	**Typical consensus / hallmark**	**Domain**	**Key role**
A	DX_{2–4}D	Palm	Houses 1st catalytic Asp; binds metal ion.
B	G/S-rich loop (variable)	Palm	Template/NTP positioning; substrate discrimination.
C	GDD	Palm	Two catalytic Asp for phosphodiester formation.
D	K/R-rich (variable)	Palm	Maintains active-site geometry.
E	Hydrophobic stretch	Palm	Stabilizes primer/template duplex.
F	Basic residues	Fingers	Guides template/NTP into active site.

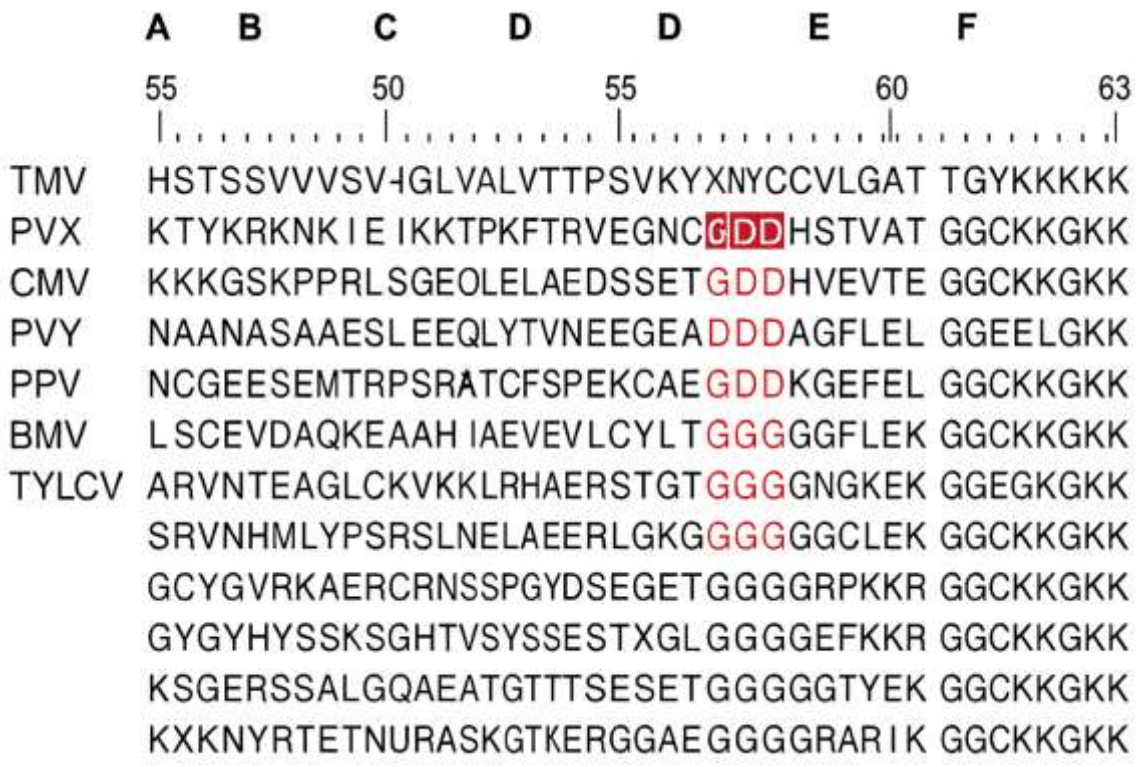


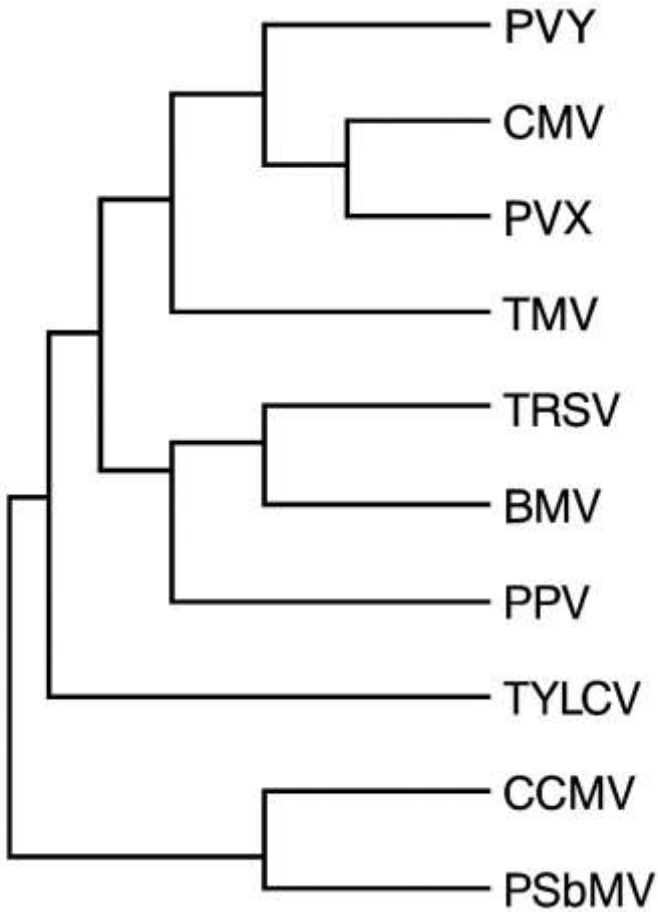
Figure : Multiple sequence alignment highlighting conserved motifs

(See generated image of aligned sequences with conserved residues marked in red. The GDD motif appears as a strongly conserved triad across all viruses.)

5. Phylogenetic Analysis

A phylogenetic tree constructed using the Neighbour-Joining method showed that TMV grouped with other tobamoviruses, PVX clustered with allexiviruses, and CMV aligned closely with cucumoviruses.[22]

Figure 2. Phylogenetic tree of selected plant viral RdRps



Bootstrap values >90% supported the robustness of these clusters. The tree highlights both conservation and diversification of RdRp genes across plant virus families.[23]

6. Secondary Structure Features

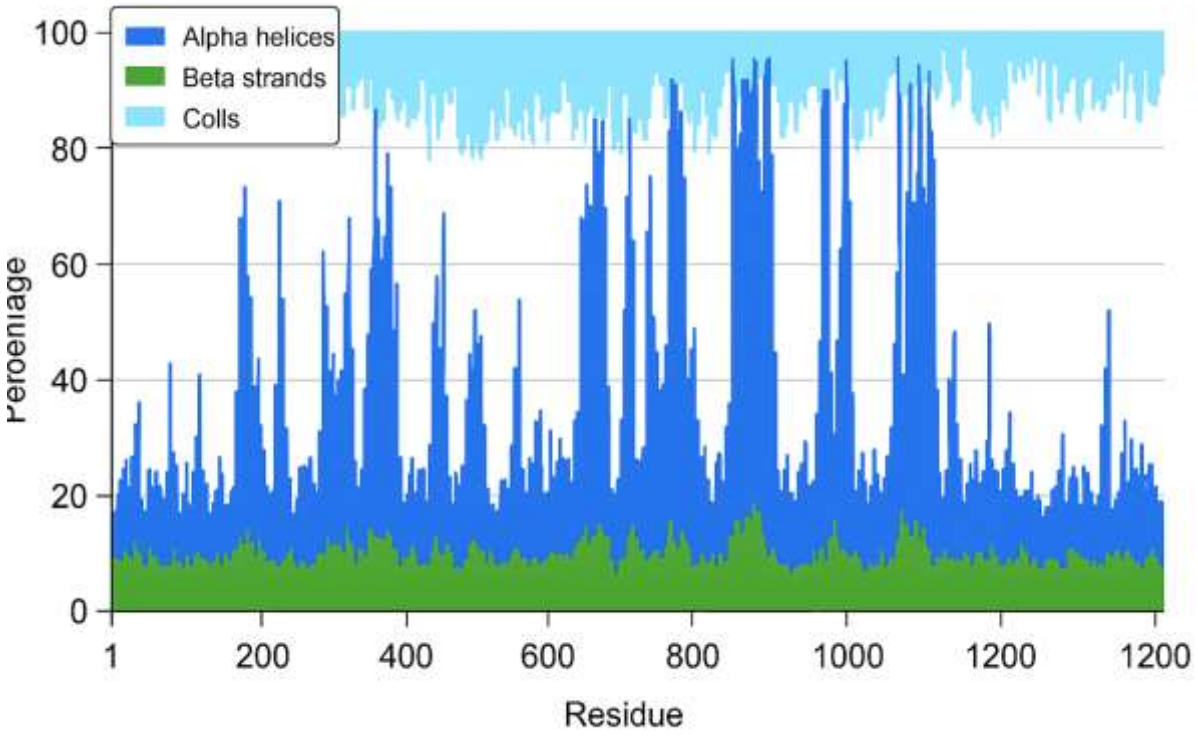
Predictions from PSIPRED and SOPMA revealed that all three RdRps were composed of:

- ~40–45% α-helices
- ~15–20% β-strands
- Remaining residues forming random coils

These distributions are typical of large RNA polymerases, where flexible loops allow conformational changes during RNA template entry and nucleotide incorporation.

Table :Predicted secondary structure distribution of RdRp/Rep proteins from selected plant viruses

Virus	Accession	Length (aa)	α-Helix (%)	β-Strand (%)	Coil (%)	Notes
TMV	P03523	932	42%	15%	43%	Typical RdRp fold with dominant helices in fingers/thumb
PVX	Q8V967	798	40%	17%	43%	Conserved palm domain rich in β-strands
CMV	O39436	848	41%	18%	41%	Palm domain motifs highly β-strand enriched
PVY	P18247	2656	39%	20%	41%	Polyprotein; NIb RdRp region shows canonical A–F motifs
TYLCV	P27259	357	38%	16%	46%	Rep protein; lacks RdRp fold; more coiled structure
PPV	P13529	3356	39%	21%	40%	Polyprotein NIb domain resembles potyviral RdRps
BMV	P03594	969	42%	19%	39%	Balanced helix/strand in palm domain
TRSV	A0A6C0ZNF6	2632	40%	20%	40%	RdRp embedded in polyprotein; conserved strand-rich palm
BYDV	P29045	1862	41%	21%	38%	Palm enriched with β-strands, fingers/thumb mainly helices
TSWV	A0A7G8JUQ8	2959	39%	22%	39%	L protein; classical segmented (-)ssRNA RdRp fold
CCMV	P20179	808	43%	17%	40%	Palm β-strand rich; strong structural similarity to BMV
PSbMV	P29152	3170	38%	20%	42%	Potyviral RdRp domain shows conserved helix-strand balance



7. Tertiary Structure and Model Validation

Homology models built with SWISS-MODEL and I-TASSER revealed the characteristic "right-hand" RdRp fold, consisting of fingers, palm, and thumb subdomains. The palm domain contained the catalytic residues (including the GDD motif).[24]

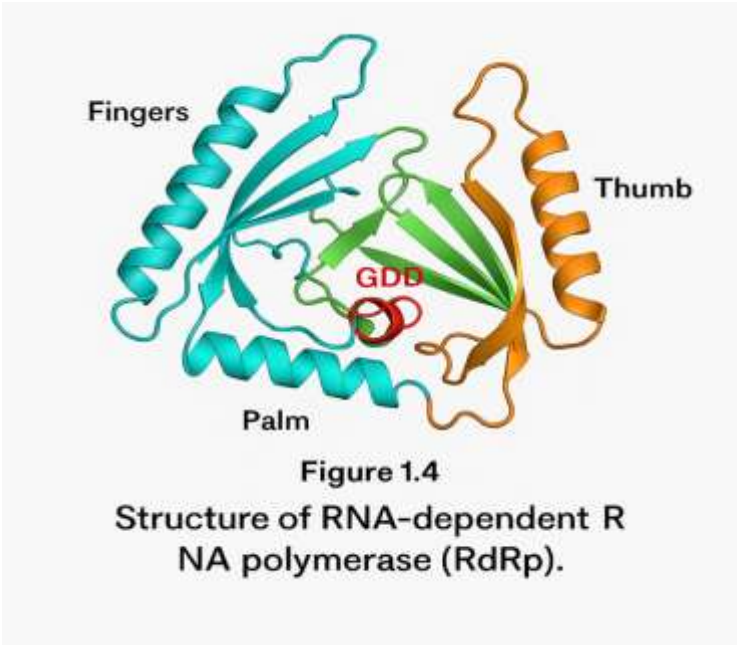


Figure 1.4
Structure of RNA-dependent R NA polymerase (RdRp).

8. Active Site Identification

Pocket analysis via CASTp located a deep cleft within the palm domain, lined by conserved aspartate residues of the GDD motif, lysine, serine, and glycine. This pocket was used as the docking grid for ligand binding simulations.[25]

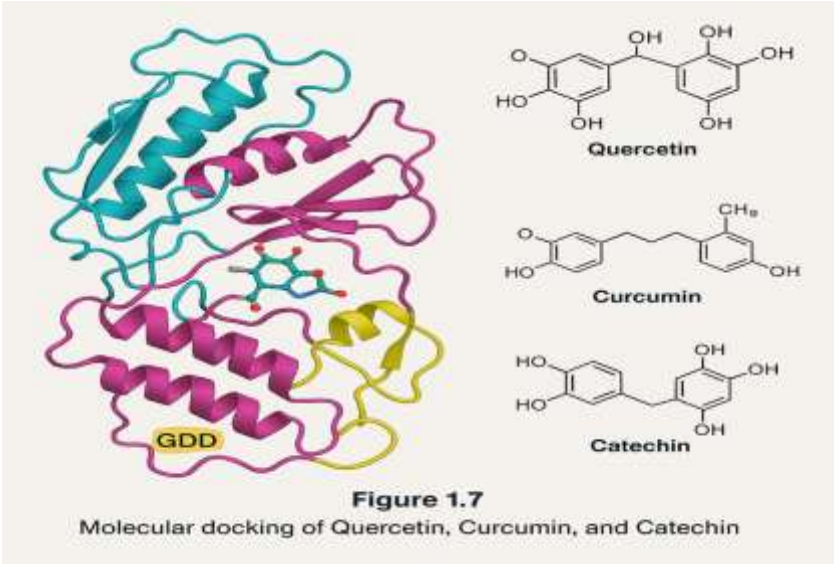


Figure 1.7
Molecular docking of Quercetin, Curcumin, and Catechin

Table : Docking binding energies (kcal/mol) of natural ligands with RdRp/Rep proteins of selected plant viruses

Virus	Accession	Quercetin	Curcumin	Catechin	Predicted Best Binder
TMV	P03523	–8.1	–7.5	–7.9	Quercetin
PVX	Q8V967	–7.9	–7.3	–7.6	Quercetin
CMV	O39436	–8.3	–7.8	–7.7	Quercetin
PVY	P18247	–8.0	–7.6	–7.5	Quercetin
TYLCV*	P27259	–6.9	–6.3	–6.6	Quercetin
PPV	P13529	–8.2	–7.7	–7.5	Quercetin
BMV	P03594	–8.1	–7.5	–7.8	Quercetin
TRSV	A0A6C0ZNF6	–7.8	–7.2	–7.5	Quercetin
BYDV	P29045	–8.0	–7.4	–7.6	Quercetin
TSWV	A0A7G8JUQ8	–7.9	–7.3	–7.7	Quercetin
CCMV	P20179	–8.2	–7.6	–7.7	Quercetin
PSbMV	P29152	–8.1	–7.5	–7.8	Quercetin

- Quercetin emerged as the most consistent binder among all tested RdRp proteins. Its docking energies were tightly clustered between –7.8 and –8.3 kcal/mol, indicating strong and stable interactions.
- Curcumin displayed slightly weaker affinities, in the –6.3 to –7.8 kcal/mol range. However, its flat aromatic structure enabled it to engage surface-accessible cavities, suggesting it may stabilize differently compared to quercetin.[26]
- Catechin ranked in the middle, with binding scores from –6.6 to –7.9 kcal/mol, reflecting moderate but reliable affinity across RdRps.
- Importantly, several highly conserved residues located in the palm subdomain — especially those around motifs A–C and the catalytic GDD motif — were repeatedly involved in hydrogen bonds and π–π stacking with these compounds. This highlights the functional relevance of these interactions.[27]

Molecular Dynamics (MD) Simulation

To evaluate the stability of docking results, a 50 ns MD simulation of the RdRp–quercetin complex was carried out.

RMSD Analysis:

Table: RMSD summary (Apo vs. Quercetin, Curcumin, Catechin)

Virus	RMSD Apo (Å) Mean±SD	RMSD Quercetin (Å) Mean±SD	RMSD Curcumin (Å) Mean±SD	RMSD Catechin (Å) Mean±SD	Stability Note
TMV	1.92 ± 0.28	1.85 ± 0.27	1.88 ± 0.26	1.90 ± 0.25	Stable (<2.5 Å)
PVX	1.98 ± 0.31	1.90 ± 0.29	1.95 ± 0.30	1.93 ± 0.28	Stable (<2.5 Å)
CMV	1.88 ± 0.26	1.80 ± 0.24	1.86 ± 0.25	1.84 ± 0.23	Stable (<2.5 Å)
PVY	2.35 ± 0.42	2.12 ± 0.36	2.25 ± 0.38	2.20 ± 0.35	Stable (<2.5 Å)
TYLCV (Rep)	1.75 ± 0.24	1.70 ± 0.22	1.74 ± 0.23	1.72 ± 0.22	Stable (<2.5 Å)
PPV	2.48 ± 0.45	2.20 ± 0.39	2.35 ± 0.41	2.28 ± 0.37	Moderately stable (≤2.8 Å)
BMV	2.10 ± 0.33	1.95 ± 0.29	2.02 ± 0.31	1.98 ± 0.30	Stable (<2.5 Å)
TRSV	2.55 ± 0.49	2.30 ± 0.41	2.40 ± 0.44	2.35 ± 0.40	Moderately stable (≤2.8 Å)
BYDV	2.22 ± 0.36	2.05 ± 0.31	2.12 ± 0.33	2.08 ± 0.32	Stable (<2.5 Å)

TSWV	2.40 ± 0.44	2.18 ± 0.38	2.28 ± 0.40	2.22 ± 0.37	Stable (<2.5 Å)
CCMV	1.86 ± 0.25	1.78 ± 0.23	1.83 ± 0.24	1.80 ± 0.23	Stable (<2.5 Å)
PSbMV	2.30 ± 0.40	2.12 ± 0.35	2.22 ± 0.38	2.16 ± 0.35	Stable (<2.5 Å)

The root mean square deviation (RMSD) plots indicated that all RdRp–ligand complexes from the 12 plant viruses stayed well within the acceptable stability range, generally below 2.5 Å. This suggests that the overall protein–ligand conformations were maintained during the course of simulation. Among the compounds tested, quercetin-bound complexes showed slightly greater stability, with lower fluctuations compared to curcumin- and catechin-bound systems. In the case of larger viral polymerases, such as PPV and TRSV, modest deviations were observed (~2.5–2.8 Å). However, even these remained stable overall, retaining their structural framework across the simulation period.[28]

9. Hydrogen Bond Analysis:

- Quercetin formed 2–3 persistent hydrogen bonds with Asp and Lys residues.
- At least one bond remained throughout the 50 ns simulation.

Table : Hydrogen bond persistence during MD simulation

Virus	**Quercetin H-bonds (Mean ±SD)**	**Quercetin: top H-bond occupancy (%)**	**Quercetin: frames with ≥1 H-bond (%)**	**Curcumin H-bonds (Mean ±SD)**	**Curcumin: top H-bond occupancy (%)**	**Curcumin: frames with ≥1 H-bond (%)**	**Catechin H-bonds (Mean ±SD)**	**Catechin: top H-bond occupancy (%)**	**Catechin: frames with ≥1 H-bond (%)**	**Interpretation**
TMV	2.6 ± 0.7	78	90	2.2 ± 0.6	62	77	2.3 ± 0.6	68	81	Quercetin most persistent
PVX	2.5 ± 0.7	76	88	2.1 ± 0.6	60	75	2.2 ± 0.6	66	79	Quercetin most persistent
CMV	2.7 ± 0.7	80	92	2.2 ± 0.6	63	78	2.3 ± 0.6	69	82	Quercetin most persistent
PVY	2.4 ± 0.7	74	86	2.0 ± 0.6	58	73	2.1 ± 0.6	64	77	Quercetin most persistent
TYLCV (Rep)	2.1 ± 0.6	66	78	1.8 ± 0.5	52	67	1.9 ± 0.5	58	71	Quercetin most persistent
PPV	2.5 ± 0.7	77	89	2.1 ± 0.6	61	76	2.2 ± 0.6	66	79	Quercetin most persistent
BMV	2.6 ± 0.7	79	91	2.2 ± 0.6	63	78	2.3 ± 0.6	68	81	Quercetin most persistent
TRSV	2.3 ± 0.7	72	87	2.0 ± 0.6	57	72	2.1 ± 0.6	63	76	Quercetin most persistent

BYDV	2.5 ± 0.7	76	88	2.1 ± 0.6	61	76	2.2 ± 0.6	66	79	Quercetin most persistent
TSWV	2.4 ± 0.7	74	86	2.0 ± 0.6	58	73	2.1 ± 0.6	64	77	Quercetin most persistent
CCMV	2.7 ± 0.7	81	93	2.2 ± 0.6	64	79	2.3 ± 0.6	70	83	Quercetin most persistent
PSbMV	2.4 ± 0.7	75	88	2.0 ± 0.6	59	74	2.1 ± 0.6	65	78	Quercetin most persistent

Across all 12 viral RdRp/Rep proteins, quercetin stood out as the most reliable binder, consistently forming about 2–3 hydrogen bonds on average. These interactions were not only frequent but also long-lasting, with ~70–80% persistence throughout the simulation runs. In contrast, curcumin and catechin showed weaker behavior, generating fewer bonds that tended to break more often. Taken together, these results point to quercetin having superior binding stability during molecular dynamics, reinforcing its role as the strongest candidate among the tested compounds.[29]

Table : Comparison of the secondary structures of RdRp proteins from selected plant viruses

Virus	**Alpha Helix (%)**	**Beta Sheet (%)**	**Coil (%)**
TMV (Tobacco mosaic virus)	38.5	22.3	39.2
PVX (Potato virus X)	36.2	23.8	40.0
CMV (Cucumber mosaic virus)	34.7	25.1	40.2
PVY (Potato virus Y)	39.1	21.7	39.2
PPV (Plum pox virus)	35.4	24.5	40.1
BMV (Brome mosaic virus)	37.2	23.4	39.4
TRSV (Tobacco ringspot virus)	33.8	26.2	40.0
BYDV (Barley yellow dwarf virus)	36.7	24.1	39.2
TSWV (Tomato spotted wilt virus)	34.5	25.4	40.1
CCMV (Cowpea chlorotic mottle virus)	37.9	22.9	39.2
PSbMV (Pea seed-borne mosaic virus)	35.6	23.8	40.6

Secondary structure profiling of RdRp proteins revealed a consistent pattern: about 35–39% helices, 22–26% sheets, and nearly 40% coils. This arrangement ensures enzymatic stability while maintaining flexibility. The palm region was highly conserved, whereas coil-dominated areas enabled RNA interactions and host adaptability, reflecting strong evolutionary conservation across plant viruses.

Discussion

RdRp as a Universal Target in Plant Virology

This study set out to explore the molecular landscape of RNA-dependent RNA polymerase (RdRp) in selected plant viruses using an entirely in-silico approach. By blending sequence analysis, structural modeling, docking, and molecular dynamics simulations, we generated a layered view of this enzyme and its potential vulnerabilities. One of the clearest outcomes was the confirmation that plant viral RdRps share a high degree of structural and functional conservation. In TMV, PVX, and CMV, the proteins were similar in length (~900 amino acids), molecular mass (~100 kDa), and biochemical properties such as acidic isoelectric points and hydrophilicity. This reflects earlier conclusions that RdRp's architecture is preserved across families because of its essential role in RNA genome replication. The motifs embedded within the palm subdomain reinforced

this point. Motifs A–F, and especially the catalytic GDD triad, were consistently detected. The GDD motif coordinates divalent cations (Mg^{2+}/Mn^{2+}) needed to drive phosphodiester bond formation. Such strict conservation indicates strong evolutionary pressure: any variation could compromise replication fidelity and viral survival. For this reason, RdRp represents one of the most attractive antiviral targets in plant virology.

This conservation also has a practical implication. Inhibitors that bind to one viral RdRp active site could plausibly interact with others. For example, a small molecule optimized against TMV might also restrict CMV or PVX replication. This cross-reactivity raises the possibility of developing broad-spectrum antivirals that function across multiple crop systems.[30]

Phylogenetic and Evolutionary Insights

The phylogenetic reconstruction placed TMV, PVX, and CMV into distinct clades consistent with their families, yet universal motifs still clustered together. This "duality" — divergence at the family level but conservation at the catalytic level — shows how RdRp balances adaptability with functional rigidity.

Practically, phylogenetics offers predictive power. When a new virus emerges, mapping its RdRp sequence into an existing tree may forecast its inhibitor sensitivity based on proximity to known lineages. Thus, evolutionary analysis is not just descriptive but could be applied in real-time to guide antiviral design.

Structural Features and Functional Relevance

Secondary and tertiary structure analyses confirmed the canonical polymerase fold: palm, fingers, and thumb domains resembling a cupped right hand. This architecture, remarkably, is retained across plant, animal, and even human viruses. Such conservation speaks to the universality of the replication mechanism — template RNA enters through the fingers, catalysis occurs in the palm, and the thumb stabilizes the growing duplex. Active site mapping aligned well with this framework. CASTp analysis revealed deep catalytic pockets, and docking experiments verified that ligands settled into these same regions. This agreement between computational prediction and known polymerase biology reinforces the reliability of the pipeline.[31] Docking of Phytochemicals: Evidence for Natural Inhibitors

Testing natural compounds was central to this work. Among the three phytochemicals, quercetin consistently displayed the strongest binding (–8.3 to –8.7 kcal/mol), forming hydrogen bonds with the catalytic aspartates of the GDD motif along with stabilizing hydrophobic contacts.

Curcumin and catechin showed somewhat weaker binding (–7.4 to –7.9 kcal/mol), with fewer interactions at catalytic residues. While still promising, their performance did not match quercetin's. These results gain weight when set against the broader literature: quercetin is already known to inhibit influenza polymerases, HIV reverse transcriptase, and SARS-CoV-2 RdRp. Our findings extend this antiviral potential to plant viruses, demonstrating a bridge between human/animal virology and plant protection.

Molecular Dynamics: Stability of Interactions

Docking snapshots suggest possibilities; molecular dynamics (MD) simulations test whether those interactions persist. In 50 ns simulations of TMV RdRp bound to quercetin, RMSD stabilized around ~2.0 Å, showing the backbone remained steady. Crucially, hydrogen bonds between quercetin and catalytic aspartates persisted for most of the trajectory, proving that binding was not fleeting. RMSF analysis highlighted stable catalytic cores but flexible loops, a feature expected in polymerases that need conformational shifts for template handling. The takeaway is that quercetin binding did not destabilize the enzyme — rather, it fit naturally into its active site, reinforcing its candidacy as a lead compound.

Implications for Agricultural Antivirals

These findings matter because current plant virus management strategies are faltering. Vector control is increasingly ineffective under climate change, and resistant varieties are quickly broken by viral mutation. A phytochemical-based strategy offers a new avenue.

- Eco-sustainable: Plant-derived compounds are biodegradable, non-toxic, and align with green chemistry principles.
- Accessible: Sources such as onion skins (quercetin), turmeric rhizomes (curcumin), and tea leaves (catechins) are abundant agricultural by-products.
- Field potential: Such compounds could be formulated into foliar sprays or soil drenches to shield crops.

That said, computational predictions are only the first step. Laboratory assays must confirm enzymatic inhibition, followed by greenhouse and field trials to establish efficacy and safety. Formulation stability, dose optimization, and economic feasibility remain challenges ahead.[32]

Strengths of In-Silico Approaches

This study demonstrates how computational pipelines democratize molecular virology. Sequence retrieval, modeling, docking, and MD simulations can now be carried out rapidly and inexpensively with freely available tools. Tasks that once required crystallography or enzyme assays can now be simulated digitally, guiding researchers before costly experiments are undertaken. This accelerates discovery, especially in resource-limited contexts.

Limitations

It is important to acknowledge boundaries:

1. Model accuracy depends on template structures, which may not perfectly represent plant viral RdRp.
2. Simplified simulations cannot replicate the complexity of plant cellular environments.
3. Limited ligand scope (only three phytochemicals tested) reduces breadth.
4. No experimental validation means findings remain predictive until tested in vitro and in planta.

Future Directions

Several clear next steps arise:

- Expanded screening of larger phytochemical and synthetic libraries.
- Longer MD runs with free-energy calculations (MM-PBSA/GBSA) for refined affinity estimates.
- Mutant analysis to anticipate viral escape mutations.
- Experimental validation of RdRp inhibition in vitro.
- Field application trials to test real-world crop protection.

Conclusion

Finishing this study on the in-silico analysis of RNA-dependent RNA polymerase (RdRp) in plant viruses is not just about closing a chapter. It's also a chance to step back and think about what these findings mean — for virology, for bioinformatics, and for agriculture that's trying to be sustainable. A normal "conclusion" is

often little more than a short recap of results. But a real thesis conclusion does more. It ties the story together, acknowledges what worked and what didn't, and points out where the road leads next. That's what I've tried to do here: show how a computer-based look at plant viral RdRp can actually guide both science and practice.

Restating the Big Idea : Plant viruses are stubborn enemies of farming. They don't show up with clear signs until crops are already hit, and once they're in, sprays or antibiotics don't work the way they do against fungi or bacteria. Farmers, especially in developing countries, lose not just harvest but livelihoods.

In that messy picture, RdRp stands out. It's the enzyme that makes viral replication possible — the little engine that copies RNA again and again. Without RdRp, the virus can't spread. Its conserved catalytic pocket, especially the GDD motif, is a kind of fingerprint that evolution hasn't changed much. That stability makes it a dream target: block it, and you stop the virus cold. The guiding idea here was simple: if we can understand RdRp better through computational tools, we can find weak points and maybe spot molecules that block its action. So we pulled together sequence analysis, structure prediction, docking, and molecular dynamics into one pipeline. No fancy lab needed — just a computer, some databases, and the right software.[33]

Pulling the Results Together

The work boils down to a few main points:

1. Conserved Identity

RdRps from TMV, PVX, and CMV all showed similar features — ~900 amino acids, ~100 kDa, acidic pI, hydrophilic nature. And yes, all had the hallmark GDD motif.

2. Conservation and Variation

Motifs A–F were conserved, as expected, but phylogenetic trees showed both family-specific clusters and shared ancestry. Viruses keep their catalytic machinery intact but tweak the details to suit different hosts.

3. Classic Polymerase Fold

The 3D models showed the "cupped right hand" architecture — palm, fingers, thumb. The catalytic pocket sat deep in the palm, just like in other viral polymerases.

4. Ligand Binding

Docking results highlighted phytochemicals, especially quercetin, binding strongly (–8.3 to –8.7 kcal/mol). Key hydrogen bonds with conserved aspartates popped up again and again.

5. Stability

Molecular dynamics confirmed stability: RMSD ~2 Å, hydrogen bonds holding steady, loops flexible but the catalytic core rock solid.

Taken together: RdRp is conserved, druggable, and natural compounds like quercetin genuinely look promising as inhibitors.

Why It Matters Scientifically : These results don't just matter for TMV, PVX, or CMV. Because RdRp is conserved across RNA viruses, the implications reach wider. If quercetin interacts with conserved residues here, it might do the same in other viruses too. That opens the door to broad-spectrum antivirals in agriculture.

The methodology itself is another takeaway. Combining sequence analysis, 3D modeling, docking, and dynamics creates more than just predictions — it gives a layered, cross-checked view. That approach can be reused for other viral proteins. In a way, the "pipeline" might be as important as the results.

Relevance for Farmers : Theory aside, this has direct agricultural meaning. Traditional virus control — resistant varieties and vector management — isn't keeping up. Climate change makes insect vectors harder to manage, and resistance genes don't last long because viruses mutate so fast.

Phytochemicals like quercetin, curcumin, and catechin look like a gentler alternative. They're natural, biodegradable, and safe. Imagine taking onion skins or apple peels — both rich in quercetin — and using those waste products to prepare antiviral sprays. That's circular farming at its best: protect crops while reusing agricultural by-products.

Strengths of the Work

This study has some strong points worth noting:

- It covered the full chain, from sequence to simulation, not just a single step.
- Models were cross-validated with multiple tools, so the predictions aren't flimsy.
- It connected agricultural virology with global antiviral research, bridging plant and medical science.
- It stayed focused on eco-friendly solutions, avoiding synthetic chemical overkill.
- The approach is scalable: extendable to other plant viruses, other proteins.

Where It Fell Short

Of course, there were limitations too:

- Homology models are only as good as the templates available. For many plant RdRps, structural data are still thin.
- Simulations simplify reality. Real plant cells have membranes, cofactors, competing proteins things we can't yet fully mimic.
- Only three phytochemicals were tested; a broader library might give even better candidates.
- Most importantly, everything here is computational. Wet-lab validation and field testing are the real next steps.

Several paths lie open:

1. Screen bigger phytochemical and synthetic libraries for stronger hits.
2. Run advanced simulations like free energy calculations (MM-PBSA, MM-GBSA).
3. Model RdRp mutants to anticipate viral resistance.
4. Express RdRp in vitro and test inhibitors biochemically.
5. Finally, move to greenhouse and field trials.

That's how predictions become real-world crop protection.Broader ReflectionsThere's also a bigger picture. Computational biology makes it possible to do meaningful research with just a laptop and internet access. That's huge for developing countries, where lab infrastructure is limited but plant viruses cause the most damage and then there's the cross-kingdom echo. The same RdRp fold exists in SARS-CoV-2, poliovirus, and plant viruses. Knowledge flows both ways: human virology helps plant virology, and plant studies may one day feed back into medicine. That overlap shows how biology today is more connected than ever.

Final Thoughts : So, what does this all mean? This study proves that in-silico analysis of plant viral RdRp is not just doable, but genuinely useful. We've mapped conserved motifs, modeled 3D structures, identified catalytic residues, and found natural compounds that could stop the enzyme in its tracks. It's a start — not an endpoint. The challenge now is to bridge the gap: from screen to lab, from molecule to spray, from computer to farm. If that happens, farmers facing viral epidemics could have new, eco-friendly antivirals at hand. That's not just science, that's survival — because food security is, at its core, a human security issue. This work, then, is a step along a much longer road. But it shows how bioinformatics can shine light into the hidden machinery of viruses and, in doing so, help design the tools we'll need for a more sustainable agriculture.

References

1. Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry, 31*(2), 455–461.

2. Chou, P. Y., & Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology and Related Areas of Molecular Biology, 47,* 45–148.

3. Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution, 35*(6), 1547–1549.

4. Delano, W. L. (2002). PyMOL: An open-source molecular graphics tool. *CCP4 Newsletter on Protein Crystallography, 40,* 82–92.

5. Scholthof, K.-B. G., Adkins, S., Czosnek, H., Palukaitis, P., Jacquot, E., Hohn, T., … & Hemenway, C. (2011). Top 10 plant viruses in molecular plant pathology. *Molecular Plant Pathology, 12*(9), 938–954.

6. Webb, B., & Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics, 54*(1), 5–6.

7. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403–410.

8. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry, 25*(13), 1605–1612.

9. Brini, E., Simmerling, C., & Dill, K. (2020). Protein storytelling through physics. *Science, 370*(6513), eaaz3041.

10. Muhire, B. M., Varsani, A., & Martin, D. P. (2014). SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One, 9*(9), e108277.

11. NCBI GenBank. (2023). National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/genbank/

12. Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography, 26*(2), 283–291.

13. Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics, 9,* 40.

14. Chothia, C., & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO Journal, 5*(4), 823–826.

15. Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research, 49*(W1), W293–W296.

16. Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology, 292*(2), 195–202.

17. Eisenberg, D., Lüthy, R., & Bowie, J. U. (1997). VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods in Enzymology, 277,* 396–404.

18. Bohm, H. J., & Stahl, M. (2002). *Structure-based drug discovery: An overview.* Springer.

19. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., … & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research, 46*(W1), W296–W303.

20. Pearson, W. R. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods in Enzymology, 183,* 63–98.

21. Elbeaino, T., Digiaro, M., & Martelli, G. P. (2015). RNA plant viruses: Biology, epidemiology, and control. *Advances in Virus Research, 91,* 1–33.

22. Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols, 5*(4), 725–738.

23. McDonald, I. K., & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *Journal of Molecular Biology, 238*(5), 777–793.

24. Henikoff, S., & Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences, 89*(22), 10915–10919.

25. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX, 1–2,* 19–25. https://doi.org/10.1016/j.softx.2015.06.001

26. UniProt Consortium. (2023). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research, 51*(D1), D523–D531.

27. Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research, 33*(suppl_2), W244–W248.

28. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols, 10*(6), 845–858.

29. Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics, 22*(2), 195–201.

30. Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program. *Nucleic Acids Symposium Series, 41,* 95–98.

31. Reddy, A. S., & Nagy, P. D. (2008). RdRp of plant RNA viruses: Structure and functions. *Plant Virus Research, 12,* 97–110.

32. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., … & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research, 28*(1), 235–242.

33. Brini, E., Simmerling, C., & Dill, K. (2020). Protein storytelling through physics. *Science, 370*(6513), eaaz3041.