



Structural, Functional Annotation of Complete Protein Coding Regions of TYLCV and Molecular Docking of Rep and Ren proteins with Resistance Genes Ty-2, Ty-5, and Ty-6 to Module Precise Molecular Auditing of Pathogen

Muhammad Tayyab¹, Sarmad Frogh Arshad^{2*}, Abdul Malik³, Muhammad Usman², Asif Saleem⁴, Imran Ahmad Khan⁵, Hasan Junaid Arshad⁶, Asma Shah Rukh⁷, Muhammad Tahir⁸, Qasim Ali Ghauri²

¹Department of Zoology, Wildlife and Fisheries, Muhammad Nawaz Shareef University of Agriculture, Multan, Pakistan; ²Department of Biochemistry and Biotechnology, Muhammad Nawaz Shareef University of Agriculture, Multan, Pakistan; ³Department of Pharmaceutics, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia; ⁴Department of Plant Breeding and Genetics, Bahauddin Zakaria University, Multan, Pakistan; ⁵Department of Pharmacy, Muhammad Nawaz Shareef University of Agriculture, Multan, Pakistan; ⁶Department of Agricultural Biochemistry and Biotechnology, Hainan University, Haikou, China; ⁷Department of Pharmacy, College of Pharmacy, Punjab University, Lahore, Pakistan; ⁸Department of Veterinary and Animal Sciences, Muhammad Nawaz Shareef University of Agriculture, Multan, Pakistan

ABSTRACT

Tomato Yellow Leaf Curl Virus (TYLCV), a single-stranded DNA begomovirus, belongs to the family *Geminiviridae*, which encodes six Open Reading Frames (ORFs). Host pathogen interaction studies revealed different resistance genes (Ty-2, Ty-5, and Ty-6) whose binding affinities and interaction patterns are still unclear against TYLCV. It is a dire need to understand the ORFs of TYLCV at the computational level to get insight into their electrostatic interaction, hydrogen bonding patterns, and binding affinities with resistance genes. In the recent study, comprehensive proteomic analyses of all encoded regions of TYLCV were performed, and the results revealed that C1 (Rep) protein was present at 30.4% in the nucleus, 34.8% in the mitochondria, and 26.1% in the cytoplasm. The 65.2% area of the mitochondria, with a 34.8% g area of the nucleus, was occupied by the C3 (Ren) factor, while the C4 encoded region showed the highest localization percentage in the mitochondria, which was 78.3% among all encoded factors. Apart from localization, the C3 (Ren) factor has a 7.94 isoelectric Pi, -0.124 hydropathicity rate, and a 28.74 instability index, indicating its stable and hydrophilic nature. Moreover, molecular docking analyses depicted that the C3 (Ren) protein had a binding affinity of -6.3 kcal/mol with the Ty-5 gene, which was the highest of all interactions, and C1 (Rep) interacted with the Ty-5 gene with a binding affinity of -6.0 kcal/mol, which encodes for replication in the host. In conclusion, results showed that C3 (Ren) was more stable and had greater binding affinity with the Ty-2, Ty-5, and Ty-6 genes in comparison to the other ORFs of TYLCV. In the future, results will be helpful in precise genome editing and hybrid peptide development to cure the infection of TYLCV.

Keywords: Tomato yellow leaf curl virus; Rep protein; Ren factor; Molecular docking; Genome editing

INTRODUCTION

Tomato (*Solanum lycopersicum*) is a flowering plant cultivated for the edible nature of fruit rich in lycopene, vitamin C, beta-

carotene, and fiber contents with several health benefits. It belongs to the family *Solanaceae* (nightshade family) and genus *Solanum*. Tomatoes originate from Western America, Central America, and Mexico and are distributed to tropical and sub-

Correspondence to: Sarmad Frogh Arshad, Department of Biochemistry and Biotechnology, Muhammad Nawaz Shareef University of Agriculture, Multan, Pakistan; E-mail: sarmad.arshad@mnsuam.edu.pk

Received: 23-Oct-2024, Manuscript No. JDMGP-24-27252; **Editor assigned:** 25-Oct-2024, PreQC No. JDMGP-24-27252 (PQ); **Reviewed:** 08-Nov-2024, QC No. JDMGP-24-27252; **Revised:** 16-Sep-2025, Manuscript No. JDMGP-24-27252 (R); **Published:** 23-Sep-2025, DOI: 10.4172/2153-0602.25.16.388

Citation: Tayyab M, Arshad SF, Malik A, Usman M, Saleem A, Khan IA, et al. (2025) Structural, Functional Annotation of Complete Protein Coding Regions of TYLCV and Molecular Docking of Rep and Ren Proteins with Resistance Genes Ty-2, Ty-5, and Ty-6 to Module Precise Molecular Auditing of Pathogen. J Data Mining Genomics Proteomics. 16:388.

Copyright: © 2025 Tayyab M, et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

tropical regions across the globe, including China, India, the United States of America, Turkey, Egypt, Italy, Iran, Brazil, and Pakistan. However, the cultivation of various crops of tomato is low due to various begomoviruses like Tomato Leaf Curl Bangalore Virus (TOLCBV), Tomato Leaf Curl New Delhi Virus (ToLCNDV), and Tomato Yellow Leaf Curl Virus (TYLCV), which is a single-stranded DNA virus that belongs to the family *Geminiviridae* and the genus *Begomovirus*.

The TYLCV is transmitted into the tomato plant by whitefly (*Bemisia tabaci*), which causes over one hundred and eleven virus species, fragmented into five strains, of which the TYLCV-Iran, Gezira, Oman, and Israeli strains are most common, but the Israeli strain of the TYLC virus appears to be the most prevalent worldwide. In terms of open reading frames, TYLCV has a monopartite genome with a genome size of about 2.8 kb, encoding six Open Reading Frames (ORFs), namely, V1, V2, C1 (Rep), C2 (TrAP), C3 (Ren), and C4. According to functional severance, V1 (CP) participates in capsid formation, V2 is involved in gene silencing, and it serves as a potent suppressor of antiviral RNAi. C1 (Rep) in the initiation of viral replication, C2 (TrAP) in transcription activation, C3 (Ren) plays a crucial role in replication enhancement, and C4 potentially contributes to symptom development and chlorosis (yellowing) of leaves [1].

Besides tomato, YLCV has been detected in many valuable crops and weed species, including sweet pepper (*Capsicum annuum* L.), chili pepper, and tobacco (*Nicotiana tabacum* L.). In tomato breeding for TYLCV resistance, the most eminent course of action is assigning virus resistance genes from wild tomato relatives to cultivated tomatoes. Recent updates have identified six resistance factors, namely *Ty-1*, *Ty-2*, *Ty-3*, *Ty-4*, *Ty-5*, and *Ty-6*, in several wild species, which have been well characterized and mapped in tomatoes. A commercial tomato hybrid called "JKHT1" was resistant to TYLCV which was created by stacking the *Ty-1/Ty-3*, *Ty-2*, *Ty-5*, and *Ty-6* genes on top of each other using marker-assisted backcross breeding. In addition, *Ty-1* and its allele *Ty-3* offer broad-spectrum resistance against various begomoviruses. Along with *Ty-1* and *Ty-3*, the *Ty-2* gene makes plants resistant to TYLCV, and *Ty-5* is especially good at fighting bipartite begomoviruses. Also, a new gene called *Ty-6* adds to the protection that *Ty-1* and *Ty-3* give against both single-partite and double-partite begomoviruses.

Furthermore, the understanding of the functional behaviors of all identified factors is vital for effective disease management and control. Although much research has been held on the functionalism of resistant genes against pathogenic activity, the *in silico* model of all resistant genes and infectious factors are still unclear. Here, we performed a comprehensive proteomic analysis of TYLCV factors and molecular docking to reveal their structural and functional behavior, paving the way for future genome editing and hybrid peptide development.

MATERIALS AND METHODS

Sequence retrieval and subcellular localization

NCBI Database was used to retrieve the FASTA sequence of all ORFs factors of TYLCV, using accession number of

(AL116122.1, CAL64778.1, CAL64777.1, BAD95504.1, ADF42089.1, BAW33041.1). Locations of all C1, C2, C3, C4, V1 and V2 factors of TYLCV in different cellular compartments were exposed through PSORT tool.

Physicochemical analysis

Various physicochemical characteristics of ORF factors of TYLCV such as molecular weight, isoelectric pi, amino acid composition, atomic composition, the total number of residues bearing positive charges and negative charges, instability index, aliphatic index, and Grand Average of Hydropathicity (GRAVY) were theoretically obtained from the ProtParam tool available at ExPASy.

Functional motif prediction

The Motif Finder tool using Pfam database were used for identification of functional sites of protein sequences.

Post translational modifications

Diverse PTM behaviors of C1, C2, C3, C4, V1 and V2 factors of TYLCV like glycosylation, phosphorylation and acetylation were determined by using Netglyc tools including NetNglyc, NetOglyc, NetCglyc, NetPhos and NetAcet respectively.

Secondary structure prediction

The PSIPRED and GOR4 web tools uncovered the structural composition and sequence plot of ORF of TYLCV.

Tertiary structure prediction

The homologous structure, multiple sequence alignment and Ramachandran plot of C1, C2, C3, C4, V1 and V2 factors of TYLCV were predicted with the help of SWISS-MODEL database.

Molecular docking

Molecular docking was performed to get insight into the binding powers of C1, and C3 factors of TYLCV with *Ty-2*, *Ty-5* and *Ty-6* genes for hybrid peptide development.

Protein preparation: The 3D structure of C1 (PDB ID: 7vg8, Uniport ID: H6WA92) and C2 (PDB ID: 112m, Uniport ID: P27260) proteins were carried out through Protein Data Bank (PDB) and made as macromolecule on Pylx tool. Before the docking process, protein was purified on Discovery Studio Visualizer 2024.

Ligand preparation: PubChem database was used to retrieve the 3D structure of TYLCV resistance factors, namely, *Ty-2* (PubChem: 381342010), *Ty-5* (PubChem: 381342012) and *Ty-6* (PubChem: 168510209) to check the binding forces with C1 (Rep) and C3 (Ren) factors of TYLCV.

Virtual screening: Open Babel, AutoDock and Vina wizard available at Pylx tool were used to minimize the ligand energy, preparation of macromolecule, selection of macromolecule and ligand to grid them for docking.

Post-docking analysis: The electrostatic interactions, hydrogen bonds, active sites, 2D structures and visualization of docking results were obtained through Discovery Studio Visualizer 2024.

RESULTS

Subcellular localization physiochemical properties

Results exposed the presence of all ORF factors of TYLCV in cellular areas in which nucleus was occupied 30.4% by C1 (Rep), 26.1% by C2, 65.2% by C3, 78.3% by C4, 60.9% by V1 and 47.8% by V2 factors of TYLCV predicted by PSORT tool [2]. Among all six factors of TYLCV, C4 and C3 (Rep) proteins showed the highest occupation rate of nucleus while presence percentage in all other cellular organelles are mentioned in Table 1. The molecular weight, instability index, negatively charged (Asp+Glu), positively charged (Asp+Glu), aliphatic index, grand average of hydropathicity (Table 2), amino acid composition (Figure 1), atomic composition (Table 3) and other physiochemical factors of C1, C2, C3, C4, V1 and V2 factors of TYLCV were annotated by Protparam available at expasy which ere pretended in (Table 2).

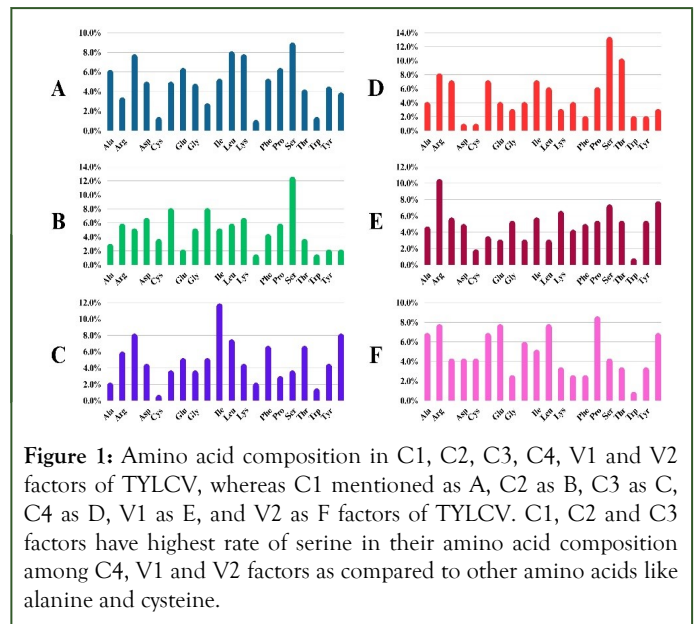


Figure 1: Amino acid composition in C1, C2, C3, C4, V1 and V2 factors of TYLCV, whereas C1 mentioned as A, C2 as B, C3 as C, C4 as D, V1 as E, and V2 as F factors of TYLCV. C1, C2 and C3 factors have highest rate of serine in their amino acid composition among C4, V1 and V2 factors as compared to other amino acids like alanine and cysteine.

Table 1: Subcellular localization of C1, C2, C3, C4, V1 and V2 factors of TYLCV.

Proteins	AN	Nucleus	Mito	Cytoplasm	Proximal	Vesicle	Peroxisome	Vacuole
C1	ALI16122.1	30.40%	34.80%	26.10%	4.30%	0%	4.30%	4.30%
C2	CAL64778.1	26.10%	69.60%	4.30%	0%	0%	0%	0%
C3	CAL64777.1	65.20%	34.80%	0%	0%	0%	0%	0%
C4	BAD95504.1	78.30%	8.70%	8.70%	0%	0%	4.30%	0%
V1	ADF42089.1	60.90%	34.80%	0%	4.30%	0%	0%	0%
V2	BAW33041.1	47.80%	47.80%	0%	0%	4.30%	0%	0%

Note: AN: Accession Number

Table 2: Presentation of theoretical pi, molecular weight, instability index and other physiochemical factors of C1, C2, C3, C4, V1 and V2 factors of TYLCV.

Proteins	AN	NAA	MW	TPi	-ve (Asp +Glu)	+ve (Asp +Glu)	Formula	TNA	II	AI	GRAVY
C1	ALI16122.1	357	40663.65	6.8	41	40	C ₁₈₂₃ H ₂₇₈ 2N ₄₉₂ O ₅₄₉ S ₉	5655	39.88	69.97	-0.665

C2	CAL6477 8.1	135	15525.29	9.08	12	17	C ₆₇₂ H ₁₀₃₂ N ₂₁₀ O ₂₀₃ S 7	2124	44.69	52.74	-0.999
C3	CAL6477 7.1	134	15924.37	7.94	13	14	C ₇₃₀ H ₁₁₂₂ N ₁₉₆ O ₁₉₇ S 4	2249	28.74	101.72	-0.124
C4	BAD9550 4.1	97	11098.54	10.5	5	11	C ₄₇₈ H ₇₆₆ N ₁₄₈ O ₁₄₇ S 5	1544	48.62	65.36	-0.72
V1	ADF4208 9.1	258	30100.56	10.09	21	44	C ₁₃₂₉ H ₂₀₈ N ₃₉₈ O ₃₇₂ S ₁₆	4197	46.95	61.9	-0.661
V2	BAW3304 1.1	116	13460.37	6.64	14	13	C ₅₉₀ H ₉₂₃ N ₁₇₅ O ₁₇₁ S 8	1867	54.42	77.33	-0.58

Note: AN: Accession Number; NAA: Number of Amino Acids; MW: Molecular Weight; TPi: Theoretical Pi; -ve: Negatively charged; +ve: Positively charged; TNA: Total Number of Atoms; II: Instability Index; AI: Aliphatic Index; GRAVY: Grand Average of Hydropathicity

Table 3: Atomic composition of C1, C2, C3, C4, V1 and V2 factors showed highest presence percentage of hydrogen as compared to carbon, nitrogen, oxygen, and sulphur.

Proteins	AN	C %	H %	N %	O %	S %
C1	ALI16122.1	32.2	48.5	8.7	9.7	0.1
C2	CAL64778.1	31.6	45.5	9.8	9.5	0.3
C3	CAL64777.1	32.4	49.8	8.7	8.7	0.1
C4	BAD95504.1	30.9	49.6	9.5	9.5	0.3
V1	ADF42089.1	31.6	49.6	9.4	8.8	0.3

Note: AN: Accession Number; C: Carbon; H: Hydrogen; N: Nitrogen; O: Oxygen; S: Sulfur

Prediction of functional motifs

Motif finder predicted 4 functional motifs in C1 (Rep), 2 in C2, 2 in C3(Ren), 1 in C4, 1 in V1 and 2 in V2 which showed various descriptions mentioned in Table 4 and Figure 2.

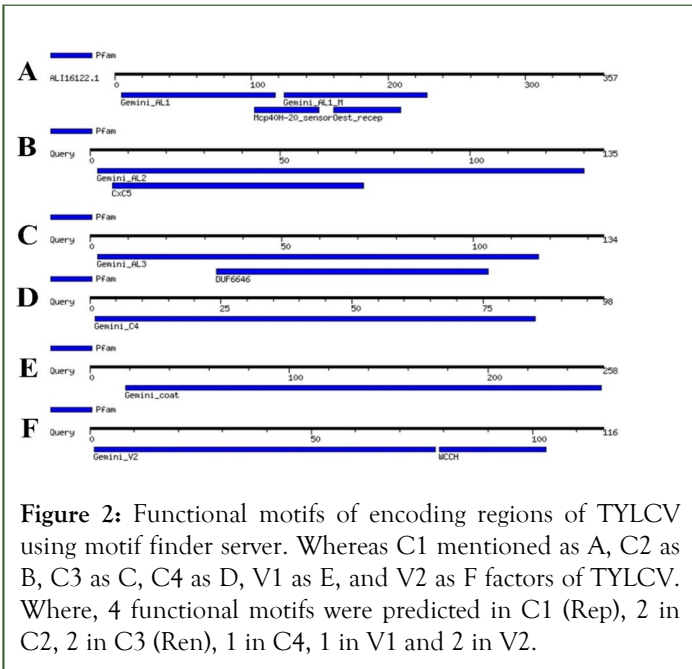


Figure 2: Functional motifs of encoding regions of TYLCV using motif finder server. Whereas C1 mentioned as A, C2 as B, C3 as C, C4 as D, V1 as E, and V2 as F factors of TYLCV. Where, 4 functional motifs were predicted in C1 (Rep), 2 in C2, 2 in C3 (Ren), 1 in C4, 1 in V1 and 2 in V2.

Post translational modifications

Among all post translational modifications, glycosylation is most dominant modification because glycoproteins take part in various cellular functions namely, protein folding, cell to cell mobility and interaction of host and pathogen. C1 and V2 has 2 N-linked glycosylation residues while C2, C3, C4 and V2 has 1 N-linked glycosylation sites predicted by NetNglyc tool (Table 5 and Figure 3). Various O-linked, C-linked glycosylation, phosphorylation, and acetylation potential residues of all ORFs of TYLCV factors were predicted by NetOglyc, NetCglyc, NetPhos and NetAcet tools showed that C1 and V1 have 25 and 28 phosphorylation sites which are highest among C2, C3, C4 and V2 factors (Table 5 and Figure 4) [3,4]. Besides glycosylation and phosphorylation, no acetylation sites were predicted in all

ORF factors of TYLCV except C3 (Ren), which showed only one acetylation site (Table 5).

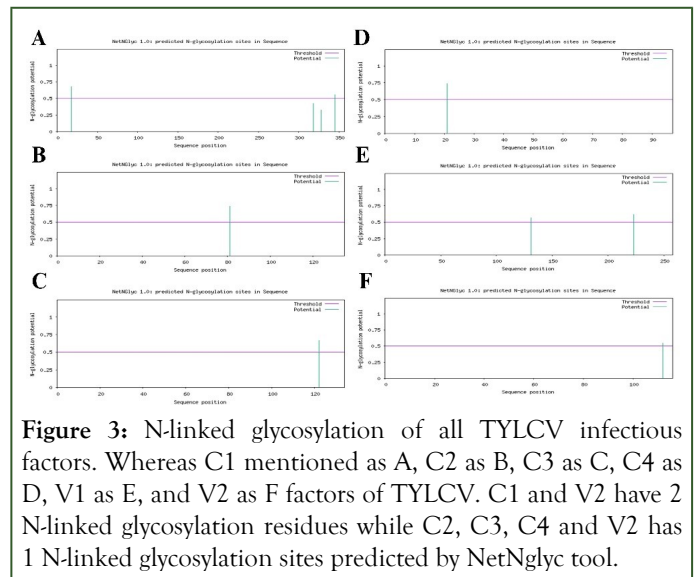


Figure 3: N-linked glycosylation of all TYLCV infectious factors. Whereas C1 mentioned as A, C2 as B, C3 as C, C4 as D, V1 as E, and V2 as F factors of TYLCV. C1 and V2 have 2 N-linked glycosylation residues while C2, C3, C4 and V2 has 1 N-linked glycosylation sites predicted by NetNglyc tool.

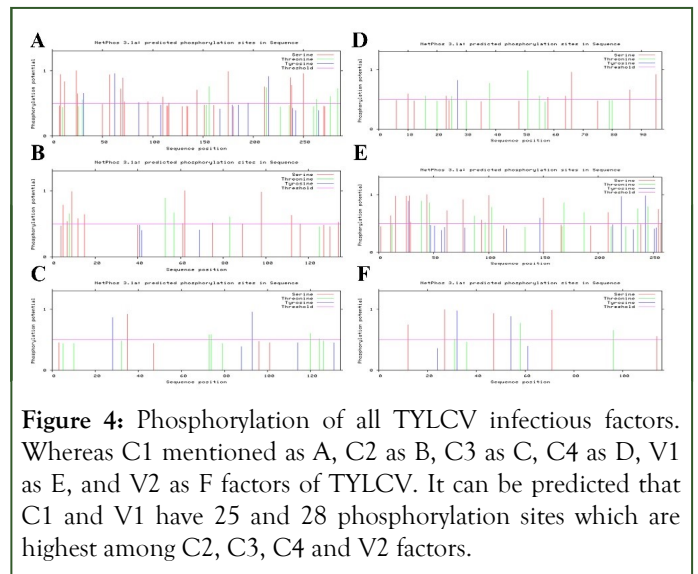


Figure 4: Phosphorylation of all TYLCV infectious factors. Whereas C1 mentioned as A, C2 as B, C3 as C, C4 as D, V1 as E, and V2 as F factors of TYLCV. It can be predicted that C1 and V1 have 25 and 28 phosphorylation sites which are highest among C2, C3, C4 and V2 factors.

Table 4: Functional motifs of all ORF factors of TYLCV.

Motifs of C1	Sequence	L	P	IE value	Description
1	FKINAKNYFLTYPNC SLSKEEALSQKLNLE TPTNKKYIKVCREP HENGEPHLHVLIQF EGKYQCKNQRFFD LSSPTRSAHFHPNIQ AAKSSTDVKTYVEK DGDFIDGVGFQID	113	5..117	4.70E-53	Gemini virus Rep catalytic domain

2	GQQSANDAYAEAL NSGSKSEALNILKEK APKDYILQFHNLNS NLDKIFQEPPAPYIPF LSSSFNQVPEELEV WVSENVMSAARP WRPNSIVIEGDSRT GKTMW	105	124..228	9.20E-36	Gemini virus rep protein central domain
3	VEKDGDFIDFGVFQ IDGRSARGGQQA NDAYAEALNSGSKS EALNILK	48	102..149	0.3	Methyl-accepting chemotaxis sensory transducer
4	FHNLNSNLDKIFQE PPAPYISPFLSSSFNQ VPEELEVWVSENV MSAARP	50	160..209	0.47	Estrogen receptor
Motifs of C2	Sequence	L	P	I-E value	Description
1	QPSSPSTSHCSQVSI KVQHKIAKKKPIRR KRVDLDCGCSYYLH LNCNNHGFTHRGT HHCSSGREWRFYL GDKQSPLFQDNRT QPAAISNEPRHHFH SDKIQPQH QEGNG DSQMFSQLPNLDDI TASDWS	129	2..130	1.10E-35	Gemini virus AL2 protein
2	PSTSHCSQVSIKVQ HKIAKKKPIRRKRV DLDCGCSYYLHLN CNNHGFTHRGTTHH CSSGREWRFYLG	67	6..72	0.045	CxC5 like cysteine cluster
Motifs of C3	Sequence	L	P	I-E value	Description
1	DSRTGELITAPQAE NGVFIWEINNPYFK ITEHSQRPFMLNHD IISIQIRFNHNIRKV MGIHKCFLNFRIWT TLQPQTGHFLRVFR YEVLYVDSLGVISI NNVIRAVDHVLYDV L	116	2..117	3.30E-43	Gemini virus AL3 protein
2	EHSQRPFMLNHDIIIS IQIRFNHNIRKVMGI HKCFLNFRIWTTLQ PQTGHFLRVFRYEV LKYVDSLGVISINN	72	33..104	0.11	Family of unknown function
Motifs of C4	Sequence	L	P	I-E value	Description
1	MGNHISMCLNSKA NTNVRTNGSSTWY PQTGQHISIRTFRQL RAQQMSRPTWRKT	85	1..85	6.50E-27	Gemini virus C4 protein

Motifs of V1	Sequence	L	P	I-E value	Description
1	ETSLILEFSKSIADQS LEEVSNLPTTHMPR RRRLNFDSPYSSRAA VPIVQGTNKRRRSWT YRPMYRKPRIYRMY RSPDVPRGCEGPCK VQSYEQRDDIKHTG IVRCVSDVTRGSGIT HRVGKRFCVKSIYFL GK VWMDENIKKQNHT NQVMFFLVRDRRPY GNSPMDFGQVFNM FDNEPSTATVKNDL RDRFQVMRKFHAT VIGGPSGMKEQALV KRFFRINSHVTYNH QEAAKYENHTENAL LLYMACTHASNPVY ATMKIRIYFYDSIS	240	18..257	3.10E-92	nuclear export factor BR1 family
Motifs of V2	Sequence	L	P	I-E value	Description
1	MWDPLLNEFPESVH GFRCMLAIKYLQSV EETYEPTLGHDLI RDLISVVRARDYVE ATRRYNHFHARLEG SPKAELRQ	78	1..78	3.20E-43	Gemini virus V2 protein
2	PIQQPCCCPHCPRH KQATIMDVQAH	25	79..103	1.20E-12	WCCH motif

Note: L: Length; P: Position

Table 5: Different post translational modification residues of C1, C2, C3, C4, V1 and V2 proteins of TYLCV.

Proteins	AN	NetPhos	NetNglyc	NetOglyc	NetCglyc	NetAcet
		Phosphorylation	N-linked glycosylation	O-linked glycosylation	C-linked glycosylation	Acetylation
C1	ALI16122.1	8S, 12S, 24S, 25S, 57S, 62Y, 68S, 72S, 86Y, 95S, 110S, 116S, 144S, 156T, 175S, 195Y, 211S, 213T, 215Y, 237S, 238S, 250S, 263T, 277T, 284T	18N, 345N	12S, 25S, 271S	None	None
C2	CAL64778.1	5S, 7S, 8T, 9S, 12S, 15S, 53T, 57T, 61S, 62S, 75S, 83T, 98S, 112S, 116S, 134S	81N	4S, 5S, 7S, 12S, 15S, 83S, 89S, 98S	None	None

C3	CAL64777.1	28Y, 35S, 73T, 74T, 122N 93Y, 120T, 124T	NONE	NONE	3S
C4	BAD95504.1	10S, 16T, 23S, 25T, 21N 27Y, 38T, 51T, 55T, 58S, 64S, 66S, 86S, 95S	16S, 23S, 24S, 25S, 35S, 38S, 41S, 51S, 75S, 79S, 86S	None	None
V1	ADF42089.1	11S, 12T,15S, 25S, 131N, 223N 27Y, 28S, 29S, 39T, 44S, 46T, 62S, 77S, 87T, 94S, 97T, 100S, 103T, 147Y, 150S, 169T, 187T, 209S, 221Y, 236T, 243Y, 245T, 255S, 257S	12S, 25S, 28S, 29S, 39S, 46S, 62S	None	None
V2	BAW33041.1	12S, 2Y, 27S, 31T, 112N 32Y, 47S, 54Y, 58T, 71S, 96T, 114S	96S	None	None

Secondary structure confirmation assay

The structural compositions of all encoded region of TYLCV were confirmed using (Accession numbers: ALI16122.1, CAL64778.1, CAL64777.1, BAD95504.1, ADF42089.1, BAW33041.1) to penetrate its functional features which showed that random coils (Purple colour) are most abundant in all ORF factors of TYLCV as compared to alpha helix (Blue colour) and extended strands (Red colour) predicted by GOR4 tool mentioned in Table 6 and Figure 5. PSIPRED tool exposed the least hydrophobic nature which resulted that all ORF region of TYLCV are hydrophilic with nearly equivalent polar and nonpolar natures (Figure 6).

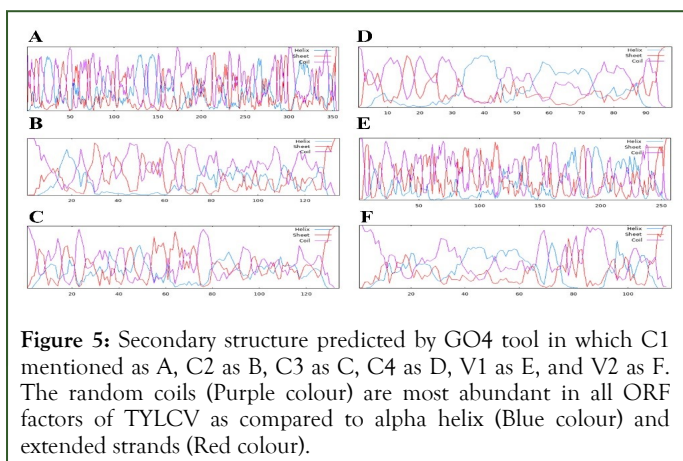
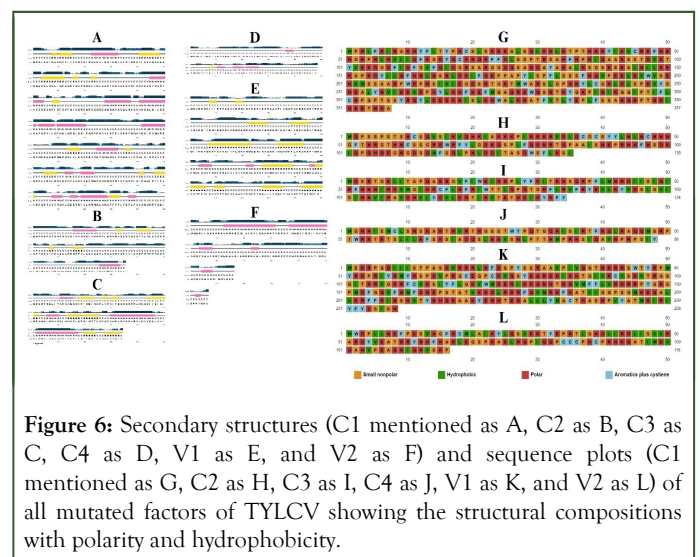


Table 6: Structural components of all infectious factors of TYLCV.

Properties	ALI16122.1	CAL64778.1	CAL64777.1	BAD95504.1	ADF42089.1	BAW33041.1
Alpha helix	20.73%	8.89%	17.91%	30.61%	21.32%	37.93%
310 helix	0%	0%	0%	0%	0%	0%
Pi helix	0%	0%	0%	0%	0%	0%
Beta bridge	0%	0%	0%	0%	0%	0%
Extended strand	26.05%	20%	30.60%	12.24%	29.46%	6.03%



Beta turn	0%	0%	0%	0%	0%	0%
Bend region	0%	0%	0%	0%	0%	0%
Random coil	53.22%	71.11%	51.49%	57.14%	49.22%	56.03%
Ambiguous states	0%	0%	0%	0%	0%	0%

Predicted 3D modelling

Using accession numbers (ALI16122.1, CAL64778.1, CAL64777.1, BAD95504.1, ADF42089.1, BAW33041.1) of all encoded factors of TYLCV, 3D modelling (C1 mentioned as A, C2 as B, C3 as C, C4 as D, V1 as E, and V2 as F in Figure 7), multiple sequence alignment (C1 mentioned as G, C2 as H, C3 as I, C4 as J, V1 as K, and V2 as L in Figure 7) and Ramachandran plots (Figure 8) were generated through SWISS-MODEL database to structurally annotate the complete ORFs of TYLCV [5].

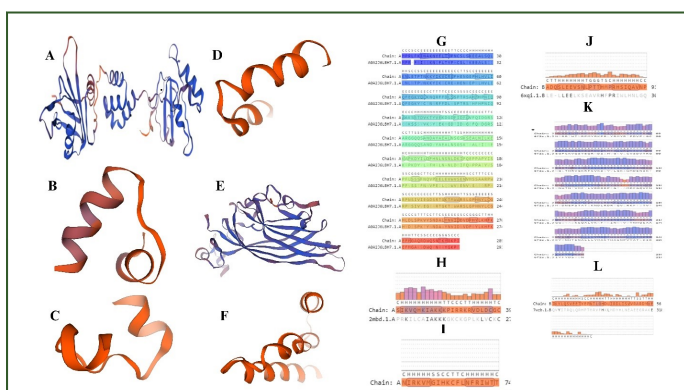


Figure 7: 3D structures of proteins C1 mentioned as A, C2 as B, C3 as C, C4 as D, V1 as E, and V2 as F. Whereas, multiple sequence alignments of proteins C1 mentioned as G, C2 as H, C3 as I, C4 as J, V1 as K, and V2 as L. Multiple sequence alignments are an excellent method for identifying interacting proteins based on sequence information and shows the protein absolute quality based on each amino acid residue.

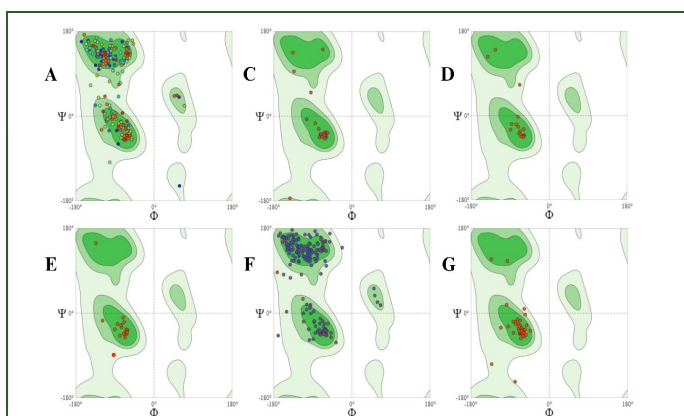


Figure 8: Graphical presentation of Ramachandran plots of all mutated factors of TYLCV. Whereas C1 mentioned as A, C2 as C, C3 as D, C4 as F, V1 as F, and V2 as G proteins of TYLCV.

Molecular docking

Results revealed that the residues of amino acids were found in the active site of C1 (Rep) comprised of various 11 amino acids with Ty-2 and Ty-5 interactions but showed 10 amino acids while interacting with Ty-6 gene (Table 7). Besides C1 factor of TYLCV, C3 (Ren) pretended to have the 7 interacting residues with Ty-2 gene, 9 with Ty-5 and 14 active site residues with Ty-6 gene which showed the highest number of interacting areas. (Table 7). Blindly docked results showed that Ty-2, Ty-5 and Ty-6 genes interacted with C1 (Rep) factor with binding affinities of -4.9 kcal/mol, -6.0 kcal/mol and -4.7 kcal/mol respectively while the C3 (Ren) protein interacted with these genes showed -5.1 kcal/mol binding affinity with Ty-2 gene, -6.3 kcal/mol with Ty-5 and -5.2 kcal/mol with Ty-6 which are nearly similar to C1 interacting rates mentioned in Table 8. Apart from binding affinities, C1 with Ty-2 gene showed 4 conventional hydrogen bonds on THR35 (1.93Å,2.17Å), THR37 (2.34Å), ASN38 (2.35Å), LYS39 (1.93Å) residue and 1 Pi-cation interaction on LYS39 (3.75Å) site which showing the novel behavior among all ORFs factors of TYLCV (Table 7). While visualization of C1 (Rep) and C3 (Ren) interactions with Ty-2, Ty-5 and Ty-6 genes are shown in Figure 9. Graphical presentation of interactions of C1 (Rep) and C3 (Ren) with all resistance genes is exposed in Figure 10 [6-8].

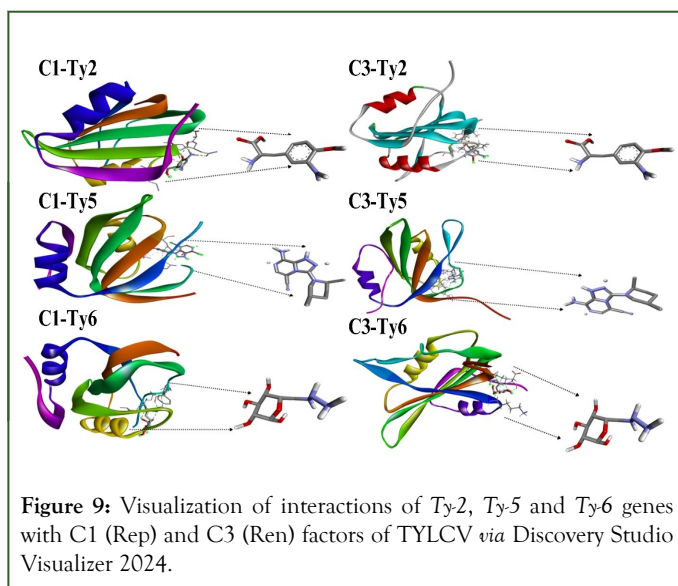


Figure 9: Visualization of interactions of Ty-2, Ty-5 and Ty-6 genes with C1 (Rep) and C3 (Ren) factors of TYLCV via Discovery Studio Visualizer 2024.

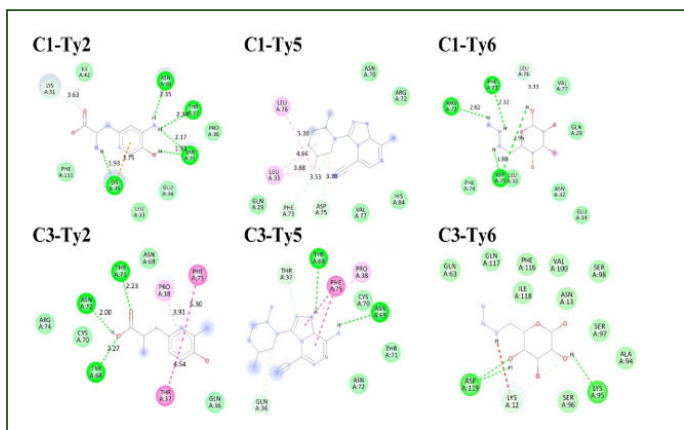


Figure 10: Structure (2D) of interactions of Ty-2, Ty-5 and Ty-6 genes with C1 (Rep) and C3 (Ren) of TYLCV via Discovery Studio Visualizer 2024. The green curves indicate hydrogen bonding, and the pink dotted lines indicate the hydrophobic interactions.

Table 7: Electrostatic interactions and hydrogen bonds of C1 and C2 protein of TYLCV while interacting with Ty-2, Ty-5, and Ty-6 genes.

IAB	C1-Ty-2	C3-Ty-2	C1-Ty-5	C3-Ty-5	C1-Ty-6	C3-Ty-6
Conventional Hydrogen bonds	THR35 (1.93Å, 2.17Å), THR37 (2.34Å), ASN38 (2.35Å), LYS39 (1.93Å)	TYR68, THR71, ASN72		TYR68 (2.44Å), ASN69 (2.60Å)	ARG72 (2.72Å), PHE73 (Å), ASP75 (1.88Å, 2.96Å)	LYS95 (2.19 Å), ASP119 (2.19 Å, 2.62 Å)
Carbon-Hydrogen bonds	LEU33, GLU34, PRO36, ILE42, PHE111		PHE73 (3.53Å), ASP75 (3.30Å)	GLN36 (3.49Å), THR37 (2.50Å)	LEU76 (3.33Å)	LYS12 (2.58Å)
Unfavorable donor-donor van der Waals	LYS31 (3.63)	GLN36, ASN69, CYS70, ARG74	GLN29, ASN32, LEU33, GLU34, PHE74, VAL77	CYS70, THR71, ASN72	GLN29, ASN32, LEU33, GLU34, PHE74, VAL77	LYS12 (2.24Å) ALA96, SER96, SER97, SER98, VAL100, PHE116, GLN117
Pi-Carbon						
Pi-Cation	LYS39 (3.75Å)					
Pi-Alkyl		PRO38	PRO38 (4.98Å)			
Pi-Sulphur						
Amide Pi-Stacked		THR37	PHR75 (3.97Å, 4.74 Å)			
Pi-Pi T- Shaped		PHE75				
Pi-Anion						
Pi-Sigma						
Alkyl		LEU33 (3.83Å, 4.64 Å), LEU76 (5.30Å)				

Note: IAB: Interactions and Bonds

Table 8: Binding affinities of C1 and C2 factors of TYLCV with Ty-2, Ty-5, and genes.

Protein-ligand	Binding affinity	Conventional hydrogen bonds	Carbon-Hydrogen bonds	van der-Waal bonds
C1-Ty2	-4.9 kcal/mol	4	5	1
C3-Ty2	-5.1 kcal/mol	3	0	4
C1-Ty5	-6.0 kcal/mol	0	2	5
C3-Ty5	-6.3 kcal/mol	2	2	3
C1-Ty6	-4.7 kcal/mol	3	1	6
C3-Ty6	-5.2 kcal/mol	2	1	10

DISCUSSION

TYLCV is a viral disease of tomato plants that is transmitted by whiteflies (*Bemisia tabaci*) and causes huge losses in tomato production. In the current study, the latest computational tools were used to explore the structural and functional behavior of all ORF factors (C1, C2, C3, C4, V1, and V2) of TYLCV. After integration into the cell, infectious proteins move towards cellular organelles like the cytoplasm, nucleus, mitochondria, Golgi apparatus, endoplasmic reticulum, peroxisome, or plasma membrane, and the C1 (Rep) protein showed the highest localization score of 26.1% in the cytoplasm (Table 1) among all encoded factors of TYLCV, which will be useful in the examination of protein function, regulation, drug development, and disease understanding. *In silico* results revealed that C1 has a 40663.65 molecular weight, keeping the dominant position of the of the C1 factor among all encoded proteins of TYLCV. Besides molecular weight, various physiochemical properties, such as instability index (39.88) and the Grand Average of Hydropathicity score, exhibit the high stable nature of the C1 factor among all other factors of TYLCV (Table 2), which are critical to performing electrophoresis.

The C1 protein of TYLCV has four N-linked glyco-motifs, which were the highest compared to all encoded regions of TYLCV, demonstrating the C1 protein's effective functional behavior. The functional and experimental expression of proteins for crystallization and isolation can be determined by exposing their physiochemical properties. Protein undergoes various modifications, in which glycosylation (attachment of carbohydrates) and phosphorylation (transferring the phosphate group of ATPs) are prevalent in nature. The results showed that the C1 (Rep) protein has 25 potential phosphorylation residues and 2 for N-linked glycosylation. The C2 (Ren) protein, on the other hand, has 7 phosphorylation sites and 1 N-like glycosylation potential area (Table 5).

Hydrogen bonds between the carbonyl oxygen and amide hydrogen atoms within the polypeptide backbone, not the side chains of the amino acids, influence the local folding patterns of the protein's secondary structure. The secondary structure of the protein reveals that C1 (Rep) has a 20.73% alpha helix, indicating a less rigid structure, and 53.22% random coils, which form the main structure of the protein and offer stability adaptation to the binding partner. This indicates that a large part of the protein has a more flexible and loose arrangement, allowing for various interactions and movements. Additionally, 26.5% extended strands, which provide some flexibility but also contribute to stability along with alpha [9]. In addition to the secondary structure, the 3D models of the C1 (Rep) and C3 (Ren) proteins, which were generated from the Protein Data Bank (PDB), included all ORF factors contributed by different bonds (Hydrogen), interactions (Pi-cation), and charges. The Ramachandran plot favored score of 95.82% indicates a high probability that the protein dihedral angles (phi and psi) fall within the energetically favorable regions. This information will be useful in further evaluating genetic variations to determine their structural impact on cellular functions.

In last, the potential residues in the active sites of the C1 factor of TYLCV were composed of 11 amino acids, of which THR35 (1.93, 2.17), THR37 (2.34), ASN38 (2.35), and LYS39 (1.93) had conventional hydrogen bonds, and LEU33, GLU34, PRO36, ILE42, and PHE11 showed carbon-hydrogen bonds with Ty-2 genes. In addition to the Ty-2 gene, Table 7 displays 10 residues interacting with the Ty-6 gene, including one carbon-hydrogen bond with LYS12. The ligand's binding affinity and the negative value are inversely proportional to each other, which indicates that a lower negative value would result in a higher binding affinity for protein. It was between -4.9 and -6.3 kcal/mol, and the Ty-5 gene had the highest binding affinity at -6.3 kcal/mol (Table 8), showing that it could attract and not be affected by the C3 (Ren) factor. The pot-screening analyses showed interactions like hydrogen bonds, electrostatic interactions, and hydrophobic bonds (Table 7). These interactions influence

binding affinity, which in turn influences the strength of these interactions.

The present study will pave the way for drug development to cure or restrict the prevalence of C1 (Rep) and C3 (Ren) proteins at the functional, expressional, and recruiting transcriptional levels, as well as cover all possible proteomic divergence among the modeled Open Reading Frames of TYLCV.

CONCLUSION

In the recent study, *Ty-2*, *Ty-5*, and *Ty-6* resistance genes demonstrated the high binding affinities with the replication factors of TYLCV. By incorporating these genes, tomato crop can be effectively protected against the devastating effect of TYLCV.

CONFLICT OF INTEREST

The Authors declare that there is no conflict of interest.

AUTHORS' CONTRIBUTION STATEMENT

MT: Writing–review and amp, editing, writing–original draft, supervision, conceptualization, investigation, and project administration; SFA: Writing–review and amp, editing, conceptualization, investigation; AM: Funding, supervision, conceptualization, investigation, and project administration; MU: Supervision, conceptualization; AS: Writing–review and investigation; IAK: Writing–review and amp, editing; HJA: Writing–review and amp, editing; ASR: Writing–review and amp, editing; MT: Review and editing; QAG: Writing–review and amp, editing.

REFERENCES

1. Abbasi A, Sufyan M, Arif MJ, Sahi ST. Effect of silicon on oviposition preference and biology of *Bemisia tabaci* (Gennadius) (Homoptera: Aleyrodidae) feeding on *Gossypium hirsutum*. *Int J Pest Manag.* 2022;68:124–134.
2. Gabol AH, Gilal AA, Rajput LB, Hajano J U D, Mangrio GQ. Distribution and damage potentials of tomato leafminer, *Tuta absoluta* (Lepidoptera: Gelechiidae), Sindh, Pakistan. *Proc Pakistan Acad Sci B Life Environ Sci.* 2023;60:283–291.
3. Adhikary T, Basak P. *In silico* approach to investigate the phytochemicals of *Terminalia arjuna* as multitarget inhibitors of proteins involved with lung cancer. *Lett Drug Des Discov.* 2024;21:329–338.
4. Aoki Kinoshita KF. Functions of glycosylation and related web resources for its prediction. 2022:135–144.
5. Arshad SF, Rehana R, Saleem MA, Usman M, Arshad HJ, Rizwana R, et al. Multi epitopes vaccine design for surface glycoprotein against SARS-CoV-2 using immunoinformatic approach. *Heliyon.* 2024;10:e24186.
6. Basit A, Humza M, Majeed MZ, Shakeel M, Idrees A, Hu C, et al. Systemic resistance induced in tomato plants by *Beauveria bassiana* derived proteins against tomato yellow leaf curl virus and aphid *Myzus persicae*. *Pest Manag Sci.* 2024;80:1821–1830.
7. Bian Y, Zhang X, Chen C, Wang G, Zhang P, Liu G, et al. Establish an efficient inoculation system of tomato yellow leaf curl virus to assist tomato resistance breeding. *Scientia Hort.* 2024;324:112567.
8. Chaitanya T, Mary S. Tentative function and structure prediction of putative genes in the whole genome of ligninolytic bacteria *Brevibacillus laterosporus* DSM 25 for better elucidation of cellular metabolism. 2024;170004.
9. Chinedu SN, Bella Omunagbe M, Okafor E, Afolabi R, Adebisi E. Computational studies on 6 pyruvoyl tetrahydropterin synthase (6 PTPS) of *Plasmodium falciparum*. *Bioinform Biol Insights.* 2024;18.