Conformational behavior of coat protein in plants and association with coat protein-mediated resistance against TMV

Jatin Sharma^{1,2} · Rituraj Purohit^{1,2,3} · Vipin Hallan^{2,3}

Received: 20 August 2019 / Accepted: 7 January 2020 / Published online: 13 January 2020
© Sociedade Brasileira de Microbiologia 2020

Abstract

Tobacco mosaic virus (TMV) coat protein (CP) self assembles in viral RNA deprived transgenic plants to form aggregates based on the physical conditions of the environment. Transgenic plants in which these aggregates are developed show resistance toward infection by TMV referred to as CP-MR. This phenomenon has been extensively used to protect transgenic plants against viral diseases. The mutants T42W and E50Q CP confer enhanced CP-MR as compared to the WT CP. The aggregates, when examined, show the presence of helical discs in the case of WT CP; on the other hand, mutants show the presence of highly stable non-helical long rods. These aggregates interfere with the accumulation of MP as well as with the disassembly of TMV in plant cells. Here, we explored an atomic level insight to the process of CP-MR through MD simulations. The subunit-subunit interactions were assessed with the help of MM-PBSA calculations. Moreover, classification of secondary structure elements of the protein also provided unambiguous information about the conformational changes occurring in the two chains, which indicated toward increased flexibility of the mutant protein and seconded the other results of simulations. Our finding indicates the essential structural changes caused by the mutation in CP subunits, which are critically responsible for CP-MR and provides an *in silico* insight into the effects of these transitions over CP-MR. These results could further be utilized to design TMV-CP-based small peptides that would be able to provide appropriate protection against TMV infection.

Keywords Coat protein-mediated resistance · Coat protein · Tobacco mosaic virus · Flexibility

Abbreviations

CP-MR	Coat protein-mediated resistance
TMV	Tobacco mosaic virus
CP	Coat protein
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation

Rg	Radius of gyration
MD	Molecular Dynamics

Introduction

Tobacco mosaic virus is a single-stranded RNA virus belonging to the *Tobamovirus* genus, which infects the plants hailing from the *Solanaceae* family, especially the Tobacco plant [1]. TMV genome encodes for four proteins, which include two replicase proteins of size 183 and 126 kDa, one cell-to-cell movement protein (MP) of size 30 kDa, and one coat protein (CP) of size 17 kDa [1–9]. The two replicase proteins are mainly involved in the efficient replication of the viral genome, and the movement protein allows the genome to move from one cell to another [6–8, 10, 11]. All these proteins collectively form the virus replicating complexes (VRC), which are structures that assist in replication and systemic spread of the viral RNA [3, 12]. Apart from these, the coat protein is known for exclusively affecting the production/accumulation

Responsible Editor: Giliane Souza Trindade.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s42770-020-00225-0>) contains supplementary material, which is available to authorized users.

✉ Rituraj Purohit
rituraj@ihbt.res.in

- ¹ Structural Bioinformatics Lab, CSIR-Institute of Himalayan Bioresource Technology (CSIR-IHBT), Palampur, HP 176061, India
- ² Biotechnology division, CSIR-IHBT, Palampur, HP 176061, India
- ³ Academy of Scientific & Innovative Research (AcSIR), CSIR-IHBT Campus, Palampur, HP 176061, India

of MP in the plant cell and plays a significant role in the fast long-distance transfer of TMV throughout the plant [5, 13–16]. The TMV, while infecting the Tobacco plant (*Nicotiana tabacum*), enters the epidermal and mesophyll cells of its leaves [5]. Once the highly stable virus is inside, it disassembles instantly due to events of virion destabilization that involves repulsion between the carboxyl group of glutamic acid present on 50th position and the carboxylate groups of aspartic acid present on the 77th position, located on the adjoining CP subunits [2, 17, 18]. With the disassembly of the virion, the initial replication cycle of the virus starts [5]. From the infected cells, the viral progeny then moves to the neighboring uninfected cells and continues to spread the infection with the help of MP [5, 19]. Long-distance movement of the viral progeny occurs once it enters the vascular tissues of the plant, and eventually, the infection spreads systemically [5]. The systemic spread of the virus is assisted by the CP [15, 20–25].

Transgenic plants, in which the CP gene from the TMV has been transferred, elicit resistance toward the infection by TMV [1, 13, 26, 27]. This concept of protection against a viral disease through the expression of a nucleotide sequence derived from a plant virus is referred to as pathogen-derived resistance (PDR) [27–30]. An extensive study on the molecular assembly of various interactions has been undertaken to construe the host–virus relationship during the resistance of infection by TMV, thereby comprehending the biophysical and conformational aspects of TMV-CP [2]. Every CP molecule is a homodimer and comprises of two identical chains, namely, chain A and chain B [31]. These CP molecules are arranged in a helical formation protecting the viral genome at the center of the helical structure [7, 9, 31–33]. The subunits in the CP molecules interact laterally and axially with each other via salt bridges and H bonds to maintain the helical structure [7, 31, 32, 34]. Resistance against the disease occurs only when the CP gene has expressed itself in the plant cells to form aggregates of the coat protein and is referred to as coat protein-mediated resistance (CP-MR) [1, 13, 27, 35]. For CP-MR to occur, the accumulation of CP in the transgenic plant is indispensable [1, 27, 35] emphasizing the fact that the protection mechanism is dependent on the CP molecules and not on CP mRNA [36, 37], also signifying the correlation between aggregation of CP molecules and CP-MR [13].

Self-assembly of CP in the absence of viral RNA depends on the pH and the ionic strength of the medium [31, 35, 38, 39]. Molecular mechanisms involved in CP-MR is different for different viruses, which are not fully understood [4, 27]. In TMV, these CP aggregates impede the disassembly of virion in the plant cell and also affects the accumulation/production of MP, preventing the cell to cell movement of the virus [4, 13, 27, 40–42]. In our study, we focused on three variants of CP, (i) Wild Type (WT), (ii) T42W mutant, and (iii) E50W mutant, depicted in Fig. 1. Mutation at these positions tends to modify the aggregation patterns of the CP molecules [17, 34, 35]. It is

already known from the previous study that the aggregation of CP molecules is necessary for CP-MR to occur, and the variants where this aggregation is absent fail to protect against the disease [1, 2, 13, 27]. Aggregates of CP molecules exhibit helical as well as non-helical arrangements [35]. In previously performed experiments by Bendahmane M et al. (1997) [27], non-helical or long rod-shaped protein aggregates have been observed in variants that incur enhanced CP-MR [27, 35, 43]. CP aggregates produced by the WT CP are primarily composed of helical discs, whereas the mutants T42W and E50Q produce highly stable rod-shaped non-helical protein aggregates [3, 7, 13, 35]. In the case of the mutant T42W, these rod-shaped aggregates tend to trap the virion blocking its disassembly sites and interfering with the disassembly of TMV [2, 13, 27, 40–42]. Apart from this, the WT CP has a positive effect on the accumulation of MP, while the mutant T42W has a negative effect on it [2, 13]. On the other hand, in the mutant E50Q, the removal of the repulsive forces between the carboxylate group of the adjacent subunits tends to increase the CP-CP interactions leading to the formation of rod-shaped aggregates thereby increasing the CP-MR activity [2, 44]. So we can say that the enhanced CP-MR in the mutants T42W and E50Q is primarily due to the long rod-shaped aggregates formed by the CP subunits [2, 3, 13]. Figure 2 gives a brief pictorial description of the fate of a TMV virion in a transgenic as well as the non-transgenic plant.

Even after extensive study on WT TMV-CP as well as on its mutants, still, many areas are uncovered. These mainly include structural aspects of CP and the effects of mutations on its conformation as well as on CP-MR. Insight into the molecular mechanism of CP-MR conferred by the WT, as well as the mutant CP, was our primary objective. In this study, we focused mainly on the structural perturbations taking place in the CP subunits for three variants of CP and the level of divergence between the mutants and the WT. Our workflow included, firstly, to explore how the mutation at position 42 and 50 bring about conformational transitions due to which the interaction between the two chains is either lost entirely or highly intensified and how these transitions differ from the ones observed in the WT and secondly, to understand how these conformational transitions alter the levels of CP-MR and to give an insight into it at an atomic level, with the help of molecular dynamics simulations.

Methods

Datasets

3D coordinates of coat protein of tobacco mosaic virus (PDB id 1EI7) [31] was retrieved from PDB database. The crystal

Fig. 1 Architecture of the full length TMV-CP: (i) Wild Type, (ii) Mutant T42W, and (iii) Mutant E50Q

structure was solved at a resolution of 2.4 Å. We removed all the crystallized heteroatoms from the PDB file. TMV-CP is a dimer having 158 amino acids in each monomer. Since the structure of the mutant T42W and E50Q is not available on PDB, so we created it by substituting tryptophan at 42nd position in place of threonine and glutamine at the 50th position in place of glutamic acid, originally present in the WT. The mutation was introduced in both the chains of a CP molecule. This mutant was created using discovery studio package 2017R2, where the “build mutants” protocol was followed. Ionization state and terminal residues charges for the WT and mutant proteins were built using the “prepare protein” protocol of discovery studio. Furthermore, RAMPAGE server [45] was used to produce Ramachandran plots for the WT and mutant proteins.

Molecular dynamics simulation

Molecular dynamics simulations were carried out through GPU version of GROMACS 5.0.6 [46–48] using the CHARMM27 force field [49]. All the simulations were performed on, on MPI-enabled high performance computing facility. WT and the mutant protein dimer were kept in a solvent box containing 48,076 and 48,081 water molecules, respectively, keeping the pH ~ 7.4 [40, 50, 51]. The protein supported an overall negative charge at this pH, so in order to electrically neutralize the net charge of the solvated cubic box, we added sodium ions (Na⁺) by using the “gmX genion” script. Furthermore, we applied energy minimization protocols to stabilize the system and eventually applied position restrained molecular dynamics simulation. The entire process comprised of two steps, firstly, within a canonical ensemble for 10 ns,

secondly, within a constant pressure-temperature ensemble for another 10 ns. Pressure and density equilibrium were achieved by restraining the entire system except solvent molecules. Following that, production runs were carried out in triplicate each for 100 ns, and the well-converged trajectories were selected for further analysis. To keep the temperature constantly at 310 K, we use the Berendsen thermostat [52], and the Ewald method is used for all the calculations involving electrostatic and coulombic interactions [53]. The current pH for the ionization states of all histidine residues was kept at 7.4. Furthermore, we restricted the bond lengths for H bonds by the SHAKE algorithm and trimmed the Coulomb and van der Waals interactions at 1.0 nm for a time step of 2 fs. The structure of the protein complexes was saved at every 0.5 ps, and the non-bonded pairs list was modified at every ten steps. The results were analyzed by “gmX dssp”, “gmX hbond”, “gmX mindist”, “gmX gyrate” and “gmX rms”, of GROMACS to get RMSD, the radius of gyration (Rg), distance between sub-units, and define secondary structure of proteins plot (DSSP) [46] and number of discrete H in the protein. For H bond analysis, we calculated the H bonds in the protein complex by a donor-acceptor distance less than 0.35 nm and of a donor-H-acceptor angle greater than 150°. Moreover, the Coot package [54] was applied to capture simulation snapshot structures and trajectory analysis. After finalizing the topologies, MM-PBSA calculations were carried out by utilizing the g_mmpbsa scripts of GROMACS [55], and it can be explained by the calculations of the free energy, according to the following equations:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{recep}} + G_{\text{lig}}) \quad (1)$$

Fig. 2 Pictorial representation of Coat Protein-Mediated Resistance

where, G_{complex} = evaluated total binding free energy of complex, G_{recep} = free receptor, G_{lig} = unbounded.

$$G_{\text{complex}} = E_{\text{MM}} + G_{\text{solv.}}^n \quad (2)$$

E_{MM} = molecular mechanical average energy, $G_{\text{solv.}}^n$ = free solvation energy.

$$E_{\text{MM}} = E_{\text{bond}} + E_{\text{non-bond}} (E_{\text{elec}} + E_{\text{vdw}}) \quad (3)$$

E_{bond} always considered as zero. Where $E_{\text{non-bond}}$ contains electrostatic and Van der Waal interactions.

$$G_{\text{solv.}}^n = G_{\text{polar}} + G_{\text{non-polar}} \quad (4)$$

$G_{\text{solv.}}^n$ energy calculated as eqⁿ (4) polar energy calculated by Poisson Boltzmann equation and non-polar via SASA (Solvent accessible surface area) as given in equation below:

$$G_{\text{non-polar}} = \gamma A + b \quad (5)$$

where γ is solvent's surface tension, A as SASA and, b is the parameter of fitting [56, 57]. The primary objective of this study is to investigate the dynamic behavior of WT as

compared to the mutant coat protein. In order to get an insight into the conformational transitions of all the proteins and check their three-dimensional space, we plotted Rg, SASA, hydrogen bonds of carbon- α of the system, and distance of subunits by utilizing Grace package.

Results

Firstly, the Ramachandran plots of the three protein complexes were used for validating the stereochemical quality of the 3D models of the mutant protein complexes. Supplementary Fig. 4 depicts the plots for the three protein complexes generated via the RAMPAGE server. Ramachandran plot analysis of the models of mutant T42W and E50Q shows that 91.6% of the residues of mutant T42W, whereas, 91.3% of the residues of mutant E50Q lie in the favorable region. The same analysis for the WT complex showed that 89.7% of residues of the WT complex lie in the favorable region. These results show that the models generated for the mutant protein complexes could be used for further MD simulation experiments.

The mutants T42W and E50Q CP are already known to produce highly stable rod-like aggregates in comparison to the less stable helical disc-shaped aggregates of the WT CP. We, therefore, performed various molecular dynamics simulations on the three protein complexes (WT and mutants T42W and E50Q) to thoroughly examine their conformational pattern with time. RMSD calculations for the protein backbones were performed thrice over a period of 100 ns for the WT, mutant T42W, and mutant E50Q protein complexes, and the results are provided in the supplementary Fig. 1, 2, and 3 respectively. Figure 3(i) depicts the minimal fluctuating rmsd results for every protein complex (Run2 for WT and Run1 for the mutant T42W and mutant E50Q protein complexes). It reveals that all the complexes at 0 ns have the same starting point, which was 0 nm for the WT and the mutants. Gradually, there was a rise in the values of all the protein complexes from 0 ns to 20 ns. After 20 ns, up to 100 ns, the RMSD value of the mutant T42W steadily increased, fluctuating at different rates and attaining peak values of 3.5 nm and 3.8 nm at 67 ns and 90 ns, respectively, and finally achieving the value of ~ 4.30 nm at 100 ns. The WT protein complex, on the other hand, followed a different trend. From 10 ns to 100 ns, RMSD values of this protein complex (WT) fluctuated around 0.5 nm and did not go beyond 0.8 nm and finally settled at the values of 0.74 nm at 100 ns. In the case of the mutant E50Q, the value of rmsd fluctuated at a rate similar to the WT. Initially, this value surpassed the WT at 11 ns but eventually decreased at 40 ns going below the WT and finally settling to a value of 0.58 nm (slightly less than the WT) at 100 ns. From 15 ns to 100 ns, the RMSD of all the complexes fluctuated at different rates, where the mutant T42W showed a significant

rise in RMSD as compared to the counterpart protein complexes. The mutant T42W protein is seen to fluctuate quite vigorously from 20 ns to 100 ns. From Fig. 3(i), it can be noticed that the maximum fluctuations are observed in mutant T42W as compared to the WT and mutant E50Q, where very minute fluctuations take place.

RMSF calculations were also carried out on C- α atoms of the three protein complexes, and the results are shown in Fig. 3(ii). In the rmsf graph, the x-axis represents the residue number, where the residue 1 to 158, belong to chain A and residue 201 to 358 belong to the chain B. We observe that the rmsf value of the mutant T42W and E50Q is more than the WT. The value of rmsf for the WT protein complex fluctuates around 0.4 nm, with rises at 100th residue of chain A and 300th residue of chain B, achieving values of 0.58 nm and 0.59 nm, respectively. A similar pattern of fluctuations is observed in the mutant E50Q with similar rises at 100th and 300th residue. However, the value of rmsf for the E50Q is always higher than that of the WT reaching peak values of 0.67 nm and 0.72 nm at 100th and 300th residue, respectively. The increased fluctuations observed at 100th and 300th residue is due to the change in conformation of the loop present in the region from residue 89 to 108 in chain A and residue 289 to 308 in chain B. This change in conformation of the loop is also depicted with the rmsf graph in Fig. 3(ii) itself. In the case of the mutant T42W, we observe that the value of rmsf was more than the values for the mutant E50Q and the WT protein complexes to a great extent for the entire simulation achieving values of 1.77 nm and 2.19 nm at 100th and 300th residues.

To further extend our analysis, we performed the solvent-accessible surface area (SASA) analysis of the three protein complexes. Figure 4(i) shows the results of the SASA analysis for all the protein structures. We can notice from the graph that in the case of the WT and the mutant E50Q, the SASA gradually decreases with time, whereas for the mutant T42W, SASA increases with time. At 0 ns, the initial value of SASA for the WT and mutant E50Q was ~ 190 nm², while the final value at 100 ns was ~ 177 nm² for the WT and slightly less (~ 175 nm²) for the mutant E50Q. The mutant E50Q and WT showed a similar pattern of SASA where a sudden drop in value is observed at 10 ns up to 16 ns from where the value increases for a while and then drops again to become stable and finally settle at that attained level itself. On the other hand, the mutant T42W has quite different fluctuations of SASA as compared to the WT and E50Q. Significant fluctuations were observed in the mutant T42W throughout the simulation time. The values for the mutant T42W continuously increased from ~ 187 nm² at 0 ns to ~ 195 nm² at 100 ns and always remained above the fluctuations of the counterpart protein complexes. Throughout the simulation time, the values of SASA for the mutant T42W were higher than the WT, whereas the value of SASA for the mutant E50Q was

initially slightly higher than WT, but gradually, it decreased and finally settled at a value less than the WT.

Rg analysis of mutants and WT protein complexes was additionally performed to second our RMSD, RMSF, and SASA results. Figure 4(ii) shows the Rg analysis results from the simulation for the three protein complexes, where it is apparent that the value of Rg for the mutant T42W instantly rises at 8 ns and stabilizes for some time followed by another rise at 60 ns which then increases continuously and attains a peak value of ~ 5.9 nm at 100 ns. Significant fluctuations have been observed in this mutant protein from 60 ns to 100 ns. While for the WT and mutant E50Q protein complexes, not many fluctuations were observed throughout the simulation time period. The initial and final values of Rg for the WT protein complex are almost the same, with a very slight decrease in value having an average initial value of ~ 2.5 nm and

attaining a final value of ~ 2.4 nm. On the other hand, the mutant E50Q also follows a pattern similar to the WT, having an initial value of ~ 2.5 nm and finally settling to a value of ~ 2.3 nm (slightly less than the WT). The mutant T42W has considerably higher values of Rg as compared to the WT and the mutant E50Q, throughout the simulation time period. Also, the fluctuations observed for the mutant were more frequent and vigorous during the complete simulation.

To transpire the conformational changes eventuating in the protein complex, we captured images of the WT and the mutant proteins folding at various intervals in the simulation time period, which are depicted in Fig. 5. Here it is quite distinct that the two chains of CP in case of the mutant T42W move apart from one another at 45 ns and do not come close to each other for the entire simulation time period. On the other hand, WT and mutant E50Q CP tend to maintain its structure

Fig. 3 (i) Time evolution of backbone RMSDs are shown as a function of time for the WT and the Mutant coat protein structures at 300 K. The symbol coding scheme is as follows: WT (black), mutant T42W (red), and mutant E50Q (green). (ii) Time evolution of C- α RMSFs are shown as a function of time for the WT and the Mutant coat protein structures at 300 K. The symbol coding scheme is as follows: WT (black), mutant T42W (red), and mutant E50Q (green)

Fig. 4 (i) Analysis of simulation trajectory of Solvent-accessible surface area (SASA) for the WT and mutant coat protein structures versus time at 300 K. The symbol coding scheme is as follows: WT (black), mutant T42W (red), and mutant E50Q (green). (ii) Analysis of simulation trajectory of radius of gyration (Rg) for the WT and Mutant coat protein structures versus time at 300 K. The symbol coding scheme is as follows: WT (black) mutant T42W (red), and mutant E50Q (green). (iii) Time evolution of number of intramolecular hydrogen bonds of the Chain A and the Chain B of the WT and mutant protein complexes during the simulations. The symbol coding scheme is as follows: WT (black), mutant T42W (red), and mutant E50Q (green)

throughout the simulation time period, although there is an increase in the interaction of the two chains of CP in the case of mutant E50Q (which is clear from the superimposition analysis and distance analysis performed further in the study). From Fig. 5, we can also discern that in the case of the mutant T42W CP, there is a complete alteration in the conformation of the two chains at 70 ns, which continues for the rest of the simulation time period. Significant structural perturbations are observed in the mutant T42W CP, where minor perturbations are observed in the mutant E50Q during the entire simulation time. Therefore, to check the interaction of the two chains of CP, we followed the H bond analysis and the distance analysis.

H bonds formed by the two chains in a CP molecule for the mutant and WT CP were also calculated to support our study. In Fig. 4(iii), H bonds formed between the two chains of all the protein complexes can be seen. The close proximity between subunits of WT and mutant E50Q are indicated by higher number of H bonds between them. It helped to maintain the compact structure throughout the simulation time period. The average number of H bonds in the WT and mutant E50Q complex gradually increase from ~3 H bonds at 0 ns to ~5 H bonds in WT and ~4 H bonds in the E50Q mutant at 100 ns. The mutant T42W complex, on the other hand, shows a decrease in the number of H bonds from ~3 H bonds at 0 ns to no H bonds at all. The number of H bonds in the mutant T42W complex is completely lost at 70 ns, and no H bond formation is observed beyond that point.

MD simulations were also carried out to measure the distance between two chains of the protein complexes at two axial positions (105 and 143), and the results are shown in fig. 6. Figure 6(i) shows the distance calculated between residue 105 of chain A and residue 105 of chain B. Here, it is distinctly visible that the distance between these two residues increases to a great extent after 5 ns in case of the mutant T42W, while the counterpart complexes show fluctuation in distance in a very narrow range. Another aspect to take notice of is that at the end of the simulation, the distance between the residues at position 105 is less in the case of the mutant E50Q as compared to the WT, indicating that the two chains of mutant E50Q have slightly moved closer to each other. A similar result is observed for the mutant T42W in the other position as well. Figure 6(ii) shows the distance between the

residue 143 of chain A and residue 143 of chain B. Even here, we observe that the distance between the two residues increases tremendously after 60 ns in case of T42W, while the other two protein complexes show very minute fluctuation in the distance between the two residues. This analysis indicates that the two chains move away from each other in case of the mutant T42W but, on the other hand, slightly moves closer to each other in case of the mutant E50Q, as compared to the WT. The average values of all the geometrical properties

generated during simulation for the three protein complexes are given in Table 1.

Finally, to determine the secondary structure elements of the three protein complexes, we made them go through DSSP classifications. Figure 7 shows the results of the DSSP program for the WT (Fig. 7(i)), the mutant T42W complex (Fig. 7(ii)), and the mutant E50Q complex (Fig. 7(iii)), where the residues from 1 to 158 belong to the chain A and beyond that point the residues belong to chain B. DSSP observations

Fig. 5 Static images of the WT and mutant protein complexes during the simulation time period

	Native	T42W	E50Q
0ns
15ns
35ns
45ns
60ns
70ns
85ns
95ns
100ns

Fig. 6 Analysis of simulation trajectory of WT and mutant coat protein complexes (i) Distance between the residue 105 of Chain A and residue 105 of Chain B. (ii) Distance between the residue 143 of Chain A and

residue 143 of Chain B. The symbol coding scheme is as follows: WT (black) mutant T42W (red), and mutant E50Q (green)

indicated that the β sheet, α -helix, turns, coil, and bends are observed in all the protein complexes. From Fig. 7, we observe that around residue Thr5, β -sheets are replaced by β -bridges at regular intervals in the WT as well as the mutant complexes. This phenomenon of replacement is quite frequent in the WT and occurs less frequently in the mutant E50Q, which is also visible in the zoomed out images in the figure. The next difference was observed around residue Asn25, where, in the mutant E50Q structure, 3-helices were observed in large numbers throughout the simulation, whereas in the case of the mutant T42W and WT, this appearance was not observed. The mutant E50Q structure also demonstrated an increased number of bends and coil conformation around Ala100 residue (belongs to the loop region) throughout the simulation, which might have affected the rmsf graph of this mutant. The sudden appearance of the turn conformation was also observed in the mutant E50Q at 80 ns on the residue

Ile125, which was not observed in the other two protein complexes. Chain B also showed some conformational changes. Between residue Phe210 and residue Ser215, which is depicted at the 175th position in the figure, a more significant number of 3-helices were observed in the case of the mutant E50Q as compared to the other counterpart proteins. Lastly, around residue Gly285 depicted at 250th position on the figure, we observe that in the mutant E50Q structure, β -sheets, bends, and turns were observed at 80 ns, which were not visible in the WT and the mutant T42W. Major differences in the secondary structure classification of the three protein complexes have been zoomed out for better clarity. All the zoomed out images are color coordinated showing the difference of residue 5 in red box, residue 100 in orange box, and residue 175 in black box.

The protein complexes were generated at different intervals of time during the simulation experiment to compare the configurations adopted by the complexes at a particular period of time and see the fluctuations. Figure 8 depicts these superimposed conformations for the three protein complexes at the specified periods. The protein complexes perfectly align with each other at 0 ns. Beyond this time interval, we see conformational changes start to appear at 15 ns. Here, we see that the chain B of the mutant T42W complex protein start to move away from the chain A, affecting both the helical as well as the loop region of the chain B where the loop region was inclined at an angle to the chain B of WT. On the other hand, the chain B of mutant E50Q tends to curl toward the chain A and adopts a conformation different from the WT. We only observe changes in the lower axial loop region of chain B of the mutant E50Q, while there are no significant differences in the helical region.

Table 1 Average values of the geometrical properties of simulated proteins

Geometrical properties	Coat protein of TMV		
	WT	T42W	E50Q
RMSD	~0.7 nm	~2.4 nm	~0.5 nm
RMSF	~0.27 nm	~1.25 nm	~0.32 nm
SASA	~180 nm ²	~190 nm ²	~177 nm ²
Rg	~2.5 nm	~4.0 nm	~2.3 nm
H Bonds	~5	0	~5
Distance (Residue 105)	~3.7 nm	~5.5 nm	~3.1 nm
Distance (Residue 143)	~1.68 nm	~3.5 nm	1.67 nm

Meanwhile, no drastic alterations are visible in chain A for any protein complex at 15 ns apart from a few changes in the loop region itself. Next, we see the superimposition of proteins at 35 ns. Here, we notice that the chain B of mutant T42W gradually comes back toward the chain A. Also, the loop region of chain B of the mutant E50Q tends to be in a different conformation than the WT. Beyond this period, the two chains of the mutant T42W move away from each other to a great extent and remain separated for the rest of the simulation experiment. The mutant E50Q undergoes many alterations at 45 ns, as visible in Fig. 8. We observe new bends in the helical region, and even the loop region was inclined at a different angle than the WT. Also, the loop region of the chain has an extended conformation as compared to the WT. The chain B of mutant T42W is entirely out of place and pointed in a different direction altogether. Next, at 60 ns, we again observe similar changes in the helical and loop region of the

mutant E50Q, whereas the chain B of mutant T42W has again adopted a conformation where its loop region is at $\sim 90^\circ$ to chain B of WT. At 70 ns, chain A of mutant T42W tends to move away and adopt a conformation not previously observed during the simulation. Apart from that, the mutant E50Q again shows alteration in the helical as well as the loop region of chain B, and in only the loop region of chain A. Also the chain B tends to move slightly toward chain A in E50Q. Next at 85 ns, the two chains of mutant T42W tend to remain separated, while changes in the conformation of E50Q were observed. The loop region of chain B tends to curl up at a different conformation than the WT, whereas some changes were observed in the helical region of the chain A. *Major* perturbations were observed at 95 ns, where the chain B of both the mutants T42W and E50Q move away from the original position, while the chain A was adequately aligned. Lastly, at 100 ns, we see the two chains of mutant T42W completely

Fig. 7 Time evolution of the secondary structural elements of Coat Protein at 300 k (DSSP classification) zooming out the major differences among the protein complexes. (i) WT, (ii) Mutant T42W, and (iii) Mutant E50Q

Fig. 8 Superimposed images of the WT and the mutant coat protein complexes at regular intervals during the simulation time period: (i)0 ns, (ii)15 ns, (iii)35 ns, (iv)45 ns, (v)60 ns, (vi)70 ns, (vii) 85 ns, (viii)95 ns, (ix)100 ns

separated from one another. On the other hand, in the case of mutant E50Q, the loop region of both chains tends to move closer to each other, decreasing the distance between them as compared to the WT.

Furthermore, we also performed MM-PBSA [57] calculations on the WT and the mutant proteins to evaluate their binding free energies and compare the subunit-subunit interactions. For this comparison, we first extracted the scripts of complexes for all the protein molecules from 98 ns to 100 ns. Table 2 depicts the SASA energy, van der Waals energy, electrostatic

energy, as well as the binding free energy, for two subunits of the CP molecule. From Table 2, it can be noticed that the minimum binding free energy (kJ/mol) is observed in the case of the mutant E50Q, whereas the max binding free energy is observed for the mutant T42W. As compared to the WT molecule having a binding energy of -289.599 kJ/mol, the two mutant proteins, namely, T42W and E50Q, have binding free energy of 12.657 kJ/mol and -358.534 kJ/mol, respectively. Total binding energy is the sum of electrostatic energy, van der Waals energy, and SASA energy.

Table 2 Total binding energy and its components calculated by MM-PBSA analysis for the WT and mutant proteins

S. No.	TMV coat protein	Total binding energy (kJ/mol)	SASA energy	Electrostatic energy (kJ/mol)	van der Waal energy (kJ/mol)
1.	WT	-289.599	-26.223	-85.923	-177.453
2.	T42W	12.657	0.227	12.435	-0.005
3.	E50Q	-358.534	-25.356	-134.928	-198.250

Discussion

Protein dynamics are not much affected by trivial changes in the protein structure, so any robust change in the dynamics of a protein is likely to be an effect of a substantial conformational change, which might affect the stability of the protein positively or negatively [58, 59]. Proteins acquire flexibility by such alterations in their conformation, which may involve folding of the whole protein [60–66]. Such changes can intervene with the biological functioning of the protein [60–67]. We, therefore, observed the key conformational changes during MD simulation.

Characterization of the structural nature of the three protein complexes was firstly done by performing the RMSD and RMSF calculations. The quantitative approach of RMSD and RMSF calculation provides us with the comparability of various protein structures under study. RMSD calculations were performed three times, and the results with minimal fluctuations were chosen for further analysis. In Fig. 3(i), depicting the minimal fluctuating RMSD results of the three protein complexes, we can observe a significant difference in the values of the WT and mutant CP. These differences in the RMSD values for the WT and mutant protein structures of CP indicate a divergence in their protein structures, which could affect the interaction of the two subunits of CP. We also conferred that these structural alterations in the CP may also affect its biological functioning and flexibility. Moreover, the RMSF graph of the three protein complexes indicated the increased flexibility of the mutant T42W due to the high value of RMSF. Also, the mutant E50Q had more flexibility than the WT with relatively high flexibility at the loop regions of chain A (residue 89 to 108) and Chain B (residue 289 to 308). The rate at which the average values fluctuated for the WT and mutant protein structures provided a basis for further conformational analysis.

Solvent accessibility surface area (SASA) studies give a measure of the total surface area of a biomolecule that is available for binding. The higher value of SASA for a protein depicted its expanded conformation, which implies it has more surface area for interaction, whereas a lower value indicated lesser interaction surface area as well as a compact conformation. SASA results for our protein complexes show large fluctuations in the case of the mutant T42W, whereas very fewer fluctuations for the WT and mutant E50Q. The substantial fluctuations in T42W mutant denote robust and consequential structural alterations. Our results imply that the mutant T42W has an expanded structure as compared to the WT and the mutant E50Q as it has higher values of SASA throughout the simulation time. While comparing the mutant E50Q with the WT, we noticed that the value of SASA for the E50Q mutant is lesser than the WT at the end of the simulation, indicating that the surface area of this mutant decreases for interaction. We also performed the H bond analysis to support

the results of SASA. Furthermore, Rg analysis and the DSSP classification also confirmed the changes in the conformational geometry of the mutant and the WT protein complexes.

Now, to establish the expansion of structure in the case of the mutant T42W, we undertook the radius of gyration (Rg) calculations for all the protein complexes. Rg proffers the dimensional information of the protein and can be elucidated as the root mean square distance of the mass weight of a group of atoms from their common center of mass. Results from Rg analysis implies that the conformation of the mutant T42W is more flexible and quite enlarged for the entire simulation time, while on the other hand, the WT and mutant E50Q tend to maintain a compact conformation. Frequent and significant fluctuations in the mutant T42W also indicate major conformational transitions that the protein might be encountering. Depending on the Rg analysis, we can say that the WT protein conformation is more sturdy in comparison to the mutant T42W, because of the expanded conformation, thereby improving its flexibility. The mutant E50Q has a much compact conformation than the WT as it has a slightly lesser value of Rg for most of the simulation time period. SASA and Rg analysis provided us with information about the conformational geometry of the mutants and WT TMV-CP. The respective conformation of the two protein complexes will surely affect the interaction between the two chains of the protein. The stability of a protein is also affected by the conformational behavior and flexibility of a protein as it gives rise to different binding interfaces not earlier present [58, 68–72]. Alterations in protein-protein interfaces may lead to changes in the function of the complex, so these interfaces have been considered for the development of new techniques to prevent such diseases [73, 74]. To further test the flexibility of the protein complexes, we also carried out H bond calculations later in our study and analyzed the results.

Molecular dynamics simulation offers a very realistic, secure, and uncomplicated atomic level approach toward understanding the patterns of protein folding and dynamics over a particular timescale [58, 62, 65, 68]. Nuclear magnetic resonance and x-ray crystallography techniques generate static models of the protein structures which provide detailed information about the conformation of a protein. These static models of all our protein structures are depicted in Fig. 5 at different intervals of time. From Fig. 5, it is very evident that there is an increase in distance between the two chains of the mutant T42W, which also suggests that there might be a loss in the interaction between them. Moreover, chain B of the mutant E50Q also tends to move closer to chain A indicating the increase in interaction between them.

Now, to test the interaction between the chains of CP and confirm the flexibility of the proteins, we carried out H bond analysis and distance analysis. Biological functioning and specificity in molecular recognition of the proteins depend

primarily on the accurate orientation as well as their interaction with other biomolecules, which mainly occur through H bonding [75]. H bond formation also accounts for a compact protein conformation, so we performed MD simulations to find out how H bond formation is correlated with the flexibility of the protein. Our results indicate toward the loss in the number of H bonds in the mutant T42W, which signifies the loss of contact between the two chains and suggests that the two chains have started to move apart from one another manifesting that the structure of the WT and mutant E50Q complex is more compact than the mutant complex. This additionally suggests increased flexibility of the mutant T42W in comparison to the WT structure. The increase in flexibility allows the mutant protein to make a large number of aggregates, which then efficiently interfere with the disassembly of the TMV inside plant cells by blocking all the possible sites. These aggregates of the mutant complex also interfere with the MP accumulation negatively, thereby enhancing CP-MR. Furthermore, from the distance analysis results, it is visible that the two chains of the T42W mutant move away from each other during the simulation time period. This study also provides a possible explanation based on H bond analysis for moving apart of the two chains of the T42W mutant protein complex. Comparing the H bonds of the mutant E50Q and WT provided us with similar patterns, but the distance analysis for E50Q showed that the distance between the two chains decreases eventually with time. This indicates increased interaction between the chains of the mutant E50Q as compared to the WT. The movement and conformational changes in the two chains are also clearly visible in the superimposition images shown in Fig. 8. The MM-PBSA calculations have also proved the increase and decrease in interactions in the mutant E50Q and T42W, respectively, as compared to the WT. During MM-PBSA calculations, the mutant E50Q showed minimum binding free energy among the three proteins being studied. Similarly, the mutant T42W has the maximum total binding energy, which stipulates decreased interaction.

Besides to gather more profound insight about the conformational transitions, we performed DSSP classification for all the protein complexes. This tool makes use of the H bond estimation algorithm to allocate secondary structures during simulation [76]. From the results mentioned above, about the DSSP classification, it can be emphatically said that the mutant T42W conformation is covered by β -sheets and β -bridges in areas which were earlier occupied by coils as well as bends in the WT. Since bridges and sheets are more stable than the coils and bends, thus this is the reason why mutant T42W CP produces highly stable aggregates that interfere with the disassembly of the virion. Also these secondary structures are considered to increase the flexibility of a protein and make it more prone to interactions. Similarly, mutant E50Q has been observed to be flanked with by conformations of α -helices and 3-helices, which were not present in the WT.

These helices are known to interact strongly with each other and stabilize the protein making it compact. Thus DSSP classification also supports the simulation results providing proof for the increased flexibility of the mutant complexes. A similar phenomenon has been studied by The Previllege Lab in the case of HIV-CP [77]. They brought about genetic manipulations in the CP of HIV, resulting in the formation of intercapsid disulfide bonds, thereby increasing the CP-CP interaction. Thus identifying a prime locale for an ambiguous display of peptides that promotes nanoscale assembly and capsid modification [77]. All the abovementioned *in silico* results can be biologically confirmed via organizing electron microscopic analysis.

Conclusion

Since TMV CP is a facilely manipulated protein, therefore, it provides a gateway for a molecular-based study of plant cell's response to conformational perturbations as well as the intracellular reactions related with them. CP is known to have multifunctional roles deciding the fate of a virus upon entry into the host, like replication, movement, transmission, and symptom development. The intensified response of the mutant CP against the viral infection is due to the change in aggregation patterns of the CP molecules in the plant. In our case, the shape of aggregates is affected by the conformational transitions taking place in the CP molecule. We therefore executed MD simulations of the WT and mutant protein complexes in pursuance of getting an atomic level insight about the conformational changes, which lead to the aforementioned abnormalities. Our results indicate significant transitions in the structure of CP molecule during a time period of 100 ns. The WT CP molecule is observed to stay intact throughout the simulations time period, while the mutant T42W CP shows loss of interaction between the two chains due to which the two chains tend to move away from one another. These structural transitions positively affected the flexibility of the mutant T42W allowing it to form a large number of highly stable aggregates which then interfere with the disassembly of the challenging virion by blocking its disassembly sites and decreasing the accumulation of MP. Similarly, the mutant E50Q shows strong interactions between the chains throughout the simulation time, as observed in the RMSF, Rg, distance analysis, and MM-PBSA calculations. The aggregates formed by the mutant E50Q are also highly stable and rod-shaped, but this phenomenon is due to the increased interactions between the chains. The DSSP study of the protein complexes to analyze in-depth the secondary structures present in them reveals that both the mutant CP has undergone some paramount changes in secondary structures present in the two chains. All the aforementioned changes are the possible reason for the development of highly stable non-helical rod-

shaped aggregates. Our results provide discernment about the molecular level alterations occurring because of the mutation T42W and E50Q, resulting in a change in patterns of aggregate formation, and these structural aspects could further be exploited for the development of TMV-CP-based small peptides which have the capability of producing CP-MR in plants.

Acknowledgements We gratefully acknowledge to the Director, CSIR-Institute of Himalayan Bioresource Technology, Palampur for providing the facilities to carry out this work. CSIR support in the form of projects MLP0201 and MLP0140 for these studies is highly acknowledged. This manuscript represents CSIR-IHBT communication no. 4494.

Data availability statement The data that support the findings of this study are openly available in Research Collaboratory for Structural Bioinformatics – Protein Data Bank (RCSB-PDB) at <https://www.rcsb.org/structure/1EI7>, PDB id – 1EI7.

Author contribution RP conceived of and designed the study. RP, JS analyzed and interpreted the data. JS drafted the paper and RP critically revised it for important intellectual content and and VP provided intellectual inputs. All authors gave final approval of the version to be published.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Abel PP, Nelson RS, De B et al (1986) Delay of disease development in transgenic plants that express the tobacco mosaic virus coat protein gene. *Science* (80-) 232(4751):738–743. <https://doi.org/10.1126/science.3457472>
- Bendahmane M, Chen I, Asurmendi S, Bazzini AA, Szecsi J, Beachy RN (2007) Coat protein-mediated resistance to TMV infection of *Nicotiana tabacum* involves multiple modes of interference by coat protein. *Virology*. 366(1):107–116. <https://doi.org/10.1016/j.virol.2007.03.052>
- Asurmendi S, Berg RH, Koo JC, Beachy RN (2004) Coat protein regulates formation of replication complexes during tobacco mosaic virus infection. *Proc Natl Acad Sci* 101(5):1415–1420. <https://doi.org/10.1073/pnas.0307778101>
- Bendahmane M, Beachy RN (1999) Control of Tobamovirus infections via pathogen-derived resistance. *Elsevier*:369–386. [https://doi.org/10.1016/S0065-3527\(08\)60357-7](https://doi.org/10.1016/S0065-3527(08)60357-7)
- Hilf ME, Dawson WO (1993) The tobamovirus capsid protein functions as a host-specific determinant of long-distance movement. *Virology*. 193(1):106–114. <https://doi.org/10.1006/viro.1993.1107>
- Deom CM, Oliver MJ, Beachy RN (1987) The 30-kilodalton gene product of tobacco mosaic virus potentiates virus movement. *Science* (80-) 237(4813):389–394. <https://doi.org/10.1126/science.237.4813.389>
- Culver JN (2002) Tobacco mosaic virus assembly and Disassembly: determinants in pathogenicity and resistance. *Annu Rev Phytopathol* 40:287–308. <https://doi.org/10.1146/annurev.phyto.40.120301.102400>
- Ishikawa M, Meshi T, Motoyoshi F, Takamatsu N, Okada Y (1986) In vitro mutagenesis of the putative replicase genes of tobacco mosaic virus. *Nucleic Acids Res* 14(21):8291–8305. <https://doi.org/10.1093/nar/14.21.8291>
- Ehrenfeld N, Gonzalez A, Cañón P, Medina C, Perez-Acle T, Arce-Johnson P (2008) Structure-function relationship between the tobamovirus TMV-cg coat protein and the HR-like response. *J Gen Virol* 89(3):809–817. <https://doi.org/10.1099/vir.0.83355-0>
- Meshi T, Watanabe Y, Saito T, Sugimoto A, Maeda T, Okada Y (1987) Function of the 30 kd protein of tobacco mosaic virus: involvement in cell-to-cell movement and dispensability for replication. *EMBO J* 6(9):2557–2563. <https://doi.org/10.1002/j.1460-2075.1987.tb02544.x>
- Pelham HRB (1978) Leaky UAG termination codon in tobacco mosaic virus RNA. *Nature*. 272(5652):469–471. <https://doi.org/10.1038/272469a0>
- Más P, Beachy RN (1999) Replication of tobacco mosaic virus on endoplasmic reticulum and role of the cytoskeleton and virus movement protein in intracellular distribution of viral RNA. *J Cell Biol* 147(5):945–958. <https://doi.org/10.1083/jcb.147.5.945>
- Bendahmane M, Szecsi J, Chen I, Berg RH, Beachy RN (2002) Characterization of mutant tobacco mosaic virus coat protein that interferes with virus cell-to-cell movement. *Proc Natl Acad Sci* 99(6):3645–3650. <https://doi.org/10.1073/pnas.062041499>
- Koshkina TE, Baranova EN, Usacheva EA, Zavriev SK (2003) A point mutation in the coat protein gene affects long-distance transport of the tobacco mosaic virus. *Mol Biol* 37(4):631–636. <https://doi.org/10.1023/A:1025155931497>
- Ding X, Shintaku MH, Carter SA, Nelson RS (2002) Invasion of minor veins of tobacco leaves inoculated with tobacco mosaic virus mutants defective in phloem-dependent movement. *Proc Natl Acad Sci* 93(20):11155–11160. <https://doi.org/10.1073/pnas.93.20.11155>
- Dawson WO (2007) Modifications of the tobacco mosaic virus coat protein gene affecting replication, movement, and symptomatology. *Phytopathology*. 78(6):783. <https://doi.org/10.1094/phyto-78-783>
- Culver JN, Dawson WO, Plonk K, Stubbs G (1995) Site-directed mutagenesis confirms the involvement of carboxylate groups in the disassembly of tobacco mosaic virus. *Virology*. 206(1):724–730. [https://doi.org/10.1016/S0042-6822\(95\)80096-4](https://doi.org/10.1016/S0042-6822(95)80096-4)
- Lu B, Stubbs G, Culver JN (1996) Carboxylate interactions involved in the disassembly of tobacco mosaic tobamovirus. *Virology*. 225(1):11–20. <https://doi.org/10.1006/viro.1996.0570>
- Rappaport I, Wildman SG (1957) A kinetic study of local lesion growth on *Nicotiana glutinosa* resulting from tobacco mosaic virus infection. *Virology*. 4(2):265–274. [https://doi.org/10.1016/0042-6822\(57\)90062-4](https://doi.org/10.1016/0042-6822(57)90062-4)
- Ding B, Haudenschild JS, Hull RJ, Wolf S, Beachy RN, Lucas WJ (2007) Secondary Plasmodesmata are specific sites of localization of the tobacco mosaic virus movement protein in transgenic tobacco plants. *Plant Cell* 4(8):915. <https://doi.org/10.2307/3869459>
- Lucas WJ, Gilbertson RL (2003) Plasmodesmata in relation to viral movement within leaf tissues. *Annu Rev Phytopathol* 32(1):387–415. <https://doi.org/10.1146/annurev.py.32.090194.002131>
- Chen MH, Citovsky V (2003) Systemic movement of a tobamovirus requires host cell pectin methylesterase. *Plant J* 35(3):386–392. <https://doi.org/10.1046/j.1365-3113X.2003.01818.x>
- Scholthof HB, Scholthof KBG, Kikkert M, Jackson AO (1995) Tomato bushy stunt virus spread is regulated by two nested genes that function in cell-to-cell movement and host-dependent systemic invasion. *Virology*. 213(2):425–438. <https://doi.org/10.1006/viro.1995.0015>
- Cadman CH (1962) Evidence for association of tobacco rattle virus nucleic acid with a cell component. *Nature*. 193(4810):49–52. <https://doi.org/10.1038/193049a0>
- Ryabov EV, Robinson DJ, Taliansky ME (2002) A plant virus-encoded protein facilitates long-distance movement of heterologous

- viral RNA. *Proc Natl Acad Sci* 96(4):1212–1217. <https://doi.org/10.1073/pnas.96.4.1212>
26. Beachy RN (1999) Coat-protein-mediated resistance to tobacco mosaic virus: discovery mechanisms and exploitation. Harrison BD, Wilson TMA, eds. *Philos Trans R Soc B Biol Sci* 354(1383): 659–664. <https://doi.org/10.1098/rstb.1999.0418>
 27. Bendahmane M, Fitchen JH, Zhang G, Beachy RN. Studies of coat protein-mediated resistance to tobacco mosaic tobamovirus: correlation between assembly of mutant coat proteins and resistance. *J Virol* 1997;71(10):7942–7950. <http://www.ncbi.nlm.nih.gov/pubmed/9311885> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC192152>. Accessed July 3, 2019
 28. Lomonosoff GP (2003) Pathogen-derived resistance to plant viruses. *Annu Rev Phytopathol* 33(1):323–343. <https://doi.org/10.1146/annurev.py.33.090195.001543>
 29. Sanford JC, Johnston SA (1985) The concept of parasite-derived resistance—deriving resistance genes from the parasite's own genome. Department of Horticultural Science, Cornell University. *J Theor Biol* 113:395–405. [https://doi.org/10.1016/S0022-5193\(85\)80234-4](https://doi.org/10.1016/S0022-5193(85)80234-4)
 30. Beachy RN (1997) Mechanisms and applications of transgenic plants. *Curr Opin Biotechnol* 8:215–220. [https://doi.org/10.1016/S0958-1669\(97\)80105-x](https://doi.org/10.1016/S0958-1669(97)80105-x)
 31. Bhyravbhatla B, Watowich SJ, Caspar DLD (1998) Refined atomic model of the four-layer aggregate of the tobacco mosaic virus coat protein at 2.4-Å resolution. *Biophys J* 74(1):604–615. [https://doi.org/10.1016/S0006-3495\(98\)77819-1](https://doi.org/10.1016/S0006-3495(98)77819-1)
 32. Bloomer AC, Champness JN, Bricogne G, Staden R, Klug A (1978) Protein disk of tobacco mosaic virus at 2.8 resolution showing the interactions within and between subunits. *Nature* 276(5686):362–368. <https://doi.org/10.1038/276362a0>
 33. Saito T, Yamanaka K, Okada Y (1990) Long-distance movement and viral assembly of tobacco mosaic virus mutants. *Virology* 176(2):329–336. [https://doi.org/10.1016/0042-6822\(90\)90002-9](https://doi.org/10.1016/0042-6822(90)90002-9)
 34. Namba K, Pattanege R, Stubbs G (1989) Visualization of protein RNA interaction in a virus: Refined structure of TMV at 2.9 Å resolution by X-ray fiber diffraction. *J Mol Biol* 208. [https://doi.org/10.1016/0022-2836\(89\)90391-4](https://doi.org/10.1016/0022-2836(89)90391-4)
 35. Asurmendi S, Berg R, Smith T (2007) Virology MB-, 2007 U. ScienceDirect - virology aggregation of TMV CP plays a role in CP functions and in coat-protein-mediated resistance. Elsevier. 366(1):98–106. <https://doi.org/10.1016/j.virol.2005.12.039>
 36. Powell PA, Sanders PR, Tumer N, Fraley RT, Beachy RN (1990) Protection against tobacco mosaic virus infection in transgenic plants requires accumulation of coat protein rather than coat protein RNA sequences. *Virology* 175(1):124–130. [https://doi.org/10.1016/0042-6822\(90\)90192-T](https://doi.org/10.1016/0042-6822(90)90192-T)
 37. Powell PA, Stark DM, Sanders PR, Beachy RN (1989) Protection against tobacco mosaic virus in transgenic plants that express tobacco mosaic virus antisense RNA. *Proc Natl Acad Sci* 86(18): 6949–6952. <https://doi.org/10.1073/pnas.86.18.6949>
 38. Jonathan P, Butler G, Durham ACH (1977) Tobacco mosaic virus protein aggregation and the virus assembly. *Adv Protein Chem* 31(C):187–251. [https://doi.org/10.1016/S0065-3233\(08\)60219-3](https://doi.org/10.1016/S0065-3233(08)60219-3)
 39. Caspar DLD (1964) Assembly and stability of the tobacco mosaic virus particle. *Adv Protein Chem* 18(C):37–121. [https://doi.org/10.1016/S0065-3233\(08\)60268-5](https://doi.org/10.1016/S0065-3233(08)60268-5)
 40. JC R, RN B. (1988) Resistance to TMV in transgenic plants results from interference with an early event in infection. *Virology* 166: 524. [https://doi.org/10.1016/0042-6822\(88\)90523-5](https://doi.org/10.1016/0042-6822(88)90523-5)
 41. Wu XJ, Beachy RN, Wilson TM, Shaw JG (1990) Inhibition of uncoating of tobacco mosaic virus particles in protoplasts from transgenic tobacco plants that express the viral coat protein gene [published erratum appears in *Virology* 1991 Mar;181(1):430]. *Virology* 179(Department of Plant Pathology, University of Kentucky, Lexington 40546):893–895. [https://doi.org/10.1016/0042-6822\(90\)90163-1](https://doi.org/10.1016/0042-6822(90)90163-1)
 42. Osbourn JK, Watts JW, Beachy RN, Wilson TMA (1989) Evidence that nucleocapsid disassembly and a later step in virus replication are inhibited in transgenic tobacco protoplasts expressing TMV coat protein. *Virology* 172(1):370–373. [https://doi.org/10.1016/0042-6822\(89\)90143-8](https://doi.org/10.1016/0042-6822(89)90143-8)
 43. Dore I, Weiss E, Altschuh D, Van Regenmortel MHV (1988) Visualization by electron microscopy of the location of tobacco mosaic virus epitopes reacting with monoclonal antibodies in enzyme immunoassay. *Virology* 162(2):279–289. [https://doi.org/10.1016/0042-6822\(88\)90467-9](https://doi.org/10.1016/0042-6822(88)90467-9)
 44. Lu B, Stubbs G, Culver JN (1998) Coat protein interactions involved in tobacco mosaic tobamovirus cross-protection. *Virology* 248(2):188–198. <https://doi.org/10.1006/viro.1998.9280>
 45. Lovell SC, Davis IW, Ili WBA et al (2003) Structure validation by Calpha geometry: Phi, psi and Cbeta deviation. *Proteins*. <https://doi.org/10.1002/prot.10286>
 46. Berendsen HJC, Hess B, Lindahl E, Van Der Spoel D, Mark AE, Groenhof G (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26(16):1701–1718. <https://doi.org/10.1002/jcc.20291>
 47. Hess B, Kutzner C, Van Der Spoel D, Lindahl E (2008) GRGMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4(3):435–447. <https://doi.org/10.1021/ct700301q>
 48. Páll S, Schulz R, Hess B et al (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>
 49. MacKerell AD, Banavali N, Foloppe N (2000) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* 56(4):257–265. [https://doi.org/10.1002/1097-0282\(2000\)56:4<257::AID-BIP10029>3.0.CO;2-W](https://doi.org/10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W)
 50. Chen MH, Sheng J, Hind G, Handa AK, Citovsky V (2000) Interaction between the tobacco mosaic virus movement protein and host cell pectin methylesterases is required for viral cell-to-cell movement. *EMBO J* 19(5):913–920. <https://doi.org/10.1093/emboj/19.5.913>
 51. Whenham RJ, Fraser RSS, Brown LP, Payne JA (1986) Tobacco-mosaic-virus-induced increase in abscisic-acid concentration in tobacco leaves: - intracellular location in light and dark-green areas, and relationship to symptom development. *Planta* 168(4):592–598. <https://doi.org/10.1007/BF00392281>
 52. Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81(8):3684–3690. <https://doi.org/10.1063/1.448118>
 53. Lee H, Pedersen LG, Essmann U, Darden T, Perera L, Berkowitz ML (2002) A smooth particle mesh Ewald method. *J Chem Phys* 103(19):8577–8593. <https://doi.org/10.1063/1.470117>
 54. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr* 60(12 I): 2126–2132. <https://doi.org/10.1107/S0907444904019158>
 55. Kumari R et al RK-J of chemical, 2014 undefined. g_mmpbsa a GROMACS tool for high-throughput MM-PBSA calculations. *ACS Publ*. <https://doi.org/10.1021/ci500020m>
 56. Kollman P, Massova I, Reyes C ... BK-A of chemical, 2000 undefined. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *ACS Publ*. <https://doi.org/10.1021/ar000033j>
 57. Massova I, Kollman PA (2000) Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect Drug Discov Des* 18:113–135. <https://doi.org/10.1023/A:1008763014207>
 58. Yesylevskyy SO, Kharkyanen VN, Demchenko AP (2006) The change of protein intradomain mobility on ligand binding: is it a

- commonly observed phenomenon? *Biophys J* 91(8):3002–3013. <https://doi.org/10.1529/biophysj.106.087866>
59. Demchenko A, Yesylevskyy S, Kharkyanen V (2006) Hierarchical clustering of the correlation patterns: new method of domain identification in proteins. *Biophys Chem* 119(1):84–93. <https://doi.org/10.1016/j.bpc.2005.07.004>
 60. Gopalakrishnan C, Jethi S, Kalsi N, Purohit R (2016) Biophysical aspect of Huntingtin protein during polyQ: an in Silico insight. *Cell Biochem Biophys* 74(2):129–139. <https://doi.org/10.1007/s12013-016-0728-7>
 61. Singh R, Bhardwaj V, Das P, Purohit R (November 2019) Natural analogues inhibiting selective Cyclin-dependent kinase protein isoforms: a computational prospective. *J Biomol Struct Dyn*:1–11. <https://doi.org/10.1080/07391102.2019.1696709>
 62. Rajendran V (2016) Structural analysis of oncogenic mutation of isocitrate dehydrogenase 1. *Mol BioSyst* 12(7):2276–2287. <https://doi.org/10.1039/c6mb00182c>
 63. Rajendran V, Gopalakrishnan C, Purohit R (2016) Impact of point mutation P29S in RAC1 on tumorigenesis. *Tumor Biol* 37(11):15293–15304. <https://doi.org/10.1007/s13277-016-5329-y>
 64. Bhardwaj V, Purohit R (May 2019) Computational investigation on effect of mutations in PCNA resulting in structural perturbations and inhibition of mismatch repair pathway. *J Biomol Struct Dyn*:1–12. <https://doi.org/10.1080/07391102.2019.1621210>
 65. Rajendran V, Gopalakrishnan C, Sethumadhavan R (2018) Pathological role of a point mutation (T315I) in BCR-ABL1 protein—a computational insight. *J Cell Biochem* 119(1):918–925. <https://doi.org/10.1002/jcb.26257>
 66. Bhardwaj V, Singh R, Singh P, Purohit R, Kumar S (2019) Elimination of bitter-off taste of stevioside through structure modification and computational interventions. *J Theor Biol* 110094. <https://doi.org/10.1016/j.jtbi.2019.110094>
 67. Teilum K, Olsen JG, Kragelund BB (2011) Protein stability, flexibility and function. *Biochim Biophys Acta - Proteins Proteomics* 1814(8):969–976. <https://doi.org/10.1016/j.bbapap.2010.11.005>
 68. Purohit R (2014) Role of ELA region in auto-activation of mutant KIT receptor: a molecular dynamics simulation insight. *J Biomol Struct Dyn* 32(7):1033–1046. <https://doi.org/10.1080/07391102.2013.803264>
 69. Sharma J, Bhardwaj V, Purohit R (2019) Structural perturbations due to mutation (H1047R) in Phosphoinositide-3-kinase (PI3K α) and its involvement in Oncogenesis: an in Silico insight. *ACS Omega* 4(14):15815–15823. <https://doi.org/10.1021/acsomega.9b01439>
 70. Kumar S, Sinha K, Sharma R, Purohit R, Padwad Y (2019) Phloretin and phloridzin improve insulin sensitivity and enhance glucose uptake by subverting PPAR γ /Cdk5 interaction in differentiated adipocytes. *Exp Cell Res*. <https://doi.org/10.1016/j.yexcr.2019.06.025>
 71. Tanwar G, Purohit R (2019) Gain of native conformation of Aurora a S155R mutant by small molecules. *J Cell Biochem* 120(7):11104–11114. <https://doi.org/10.1002/jcb.28387>
 72. Singh R, Bhardwaj V, Purohit R (2020) Identification of a novel binding mechanism of Quinoline based molecules with lactate dehydrogenase of *Plasmodium falciparum*. *Journal of Biomolecular Structure and Dynamics*, 1–15. <https://doi.org/10.1080/07391102.2020.1711809>
 73. Jubb H, Blundell TL, Ascher DB (2015) Flexibility and small pockets at protein-protein interfaces: new insights into druggability. *Prog Biophys Mol Biol* 119(1):2–9. <https://doi.org/10.1016/j.pbiomolbio.2015.01.009>
 74. Rask-Andersen M, Almén MS, Schiöth HB (2011) Trends in the exploitation of novel drug targets. *Nat Rev Drug Discov* 10(8):579–590. <https://doi.org/10.1038/nrd3478>
 75. Hubbard RE, Kamran HM (2010) Hydrogen bonds in proteins: role and strength. In: *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd, Chichester, UK. <https://doi.org/10.1002/9780470015902.a0003011.pub2>
 76. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 22(12):2577–2637. <https://doi.org/10.1002/bip.360221211>
 77. Servid A, Jordan P, O’Neil A, Prevelige P, Douglas T (2013) Location of the bacteriophage P22 coat protein C-terminus provides opportunities for the design of capsid-based materials. *Biomacromolecules*. <https://doi.org/10.1021/bm400796c>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.