

In Silico Investigation of Symptom Development Model Based on Coat Protein Interactions of Two Cucumber Mosaic Virus Strains

Masoud Akbarimotlagh

Tarbiat Modares University

Seyed Shahriar Arab

Tarbiat Modares University

Peter Palukaitis

Seoul Women's University

Masoud Shams-bakhsh (✉ shamsbakhsh@modares.ac.ir)

Tarbiat Modares University

Research Article

Keywords: coat protein, cucumber mosaic virus, pentameric structure, Ferredoxin I protein

Posted Date: February 11th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-154008/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Physiological and Molecular Plant Pathology on February 1st, 2022. See the published version at <https://doi.org/10.1016/j.pmpp.2022.101811>.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

***In Silico* Investigation of Symptom Development Model Based on Coat Protein Interactions
of Two Cucumber Mosaic Virus Strains**

Masoud Akbarimotlagh^a

^aPlant Pathology Department, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran
(masoud.akbari@modares.ac.ir)

Seyed Shahriar Arab^b

^bDepartment of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran,
Iran (sh.arab@modares.ac.ir)

Peter Palukaitis^c

^cDepartment of Horticultural Sciences, Seoul Women's University, Seoul 01797, South Korea
(scripath1@yahoo.co.uk)

Masoud Shams-Bakhsh^{a*}

^aPlant Pathology Department, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran
(shamsbakhsh@modares.ac.ir)

Running title: *In silico* interactions of Ferredoxin I with CMV CPs

1 Abstract:**2 Background**

3 Cucumber mosaic virus (CMV) causes serious diseases in agricultural crops worldwide,
4 predominantly in vegetable and ornamental plants. A critical interaction has been reported between
5 the plant Ferredoxin I (Fd I) protein with the coat protein (CP) of the M strain of CMV (M-CMV)
6 but not by the CP of the Q strain of CMV (Q-CMV). To ascertain the most likely symptom
7 development model based on CMV CP interacting with the plant Fd I protein, further *in silico*
8 investigation was performed on the CPs of the two strains (severe chlorosis symptoms (M-CMV)
9 and mild symptoms (Q-CMV)). The biological assembly of the CMV CP is composed of
10 hexameric and pentameric units.

11 Result

12 The results showed that there were significantly different interactions between pentameric and
13 hexameric form of M and Q-CMV CPs with Fd I. Pentameric structural forms of M-CMV CP
14 interact strongly with Fd I compared to pentameric structures of Q-CMV CP, while there were not
15 significantly different in affinity for hexameric structures between M and Q-CMV CPs.

16 Conclusion

17 Accordingly, it can be deduced that probably pentameric structures of M-CMV (higher
18 pathogenicity) CP block activation of plant Fd I protein. Therefore, it can be concluded that
19 trapping Fd I in the complex of pentameric structures can lead to inhibition of the activity of plant
20 Fd I in the photosynthetic electron transport chain leading the severe symptom development seen
21 in M-CMV infected plants but not in Q-CMV infected plants. It seems likely that any factor that
22 inhibits the interaction of CMV CPs with plant Fd I can prevent this symptom development.

23

24 **Keywords:** coat protein, cucumber mosaic virus, pentameric structure, Ferredoxin I protein

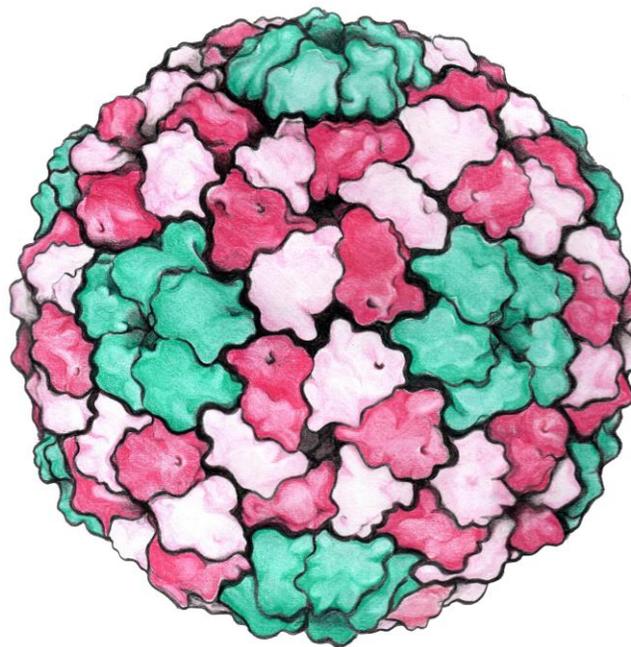
1 **1. Background:**

2
3 More than 1300 plant species are infected by cucumber mosaic virus (CMV), a member of the
4 genus *Cucumovirus* in the family *Bromoviridae* [1]. This leads to significant economic losses of
5 many crops worldwide (Scholthof et al. 2011). Members of the genus *Cucumovirus* are
6 characterized by a tripartite, plus-sense, single-stranded RNA genome that is encapsidated into
7 icosahedral particles. The genome encodes five proteins: the viral RNA replication proteins 1a
8 and 2a; the RNA-silencing suppressor protein 2b; the cell-to-cell movement protein 3a; and the
9 coat protein (CP) (Jacquemond, 2012; Palukaitis and García-Arenal, 2003; Yoon et al., 2019).

10 The symptoms of plant virus diseases may produce visible or otherwise detectable abnormalities
11 in plants. Various factors, such as the host species and variety, the physiological condition of the
12 host, age, nutrition, climate, the presence of another virus, etc., determine the nature and severity
13 of the symptoms produced. The symptoms of CMV in different plants occur as a result of complex
14 interaction between the host plant and the virus [7]. In addition some studies have shown that CP
15 and 2b interact synergistically with some other CMV components in order to induce viral
16 symptoms [8]. Many different fundamental roles occur for virus CPs in the virus infection cycle.
17 The CMV CP functions, such as the encapsidation of the viral RNAs, cell-to-cell movement,
18 systemic movement, regulation of host defense responses, aphid transmission, and symptom
19 formation have been determined [4–6, 9]. Smith et al (2000) determined the structures of CMV
20 and cowpea chlorotic mottle virus (CCMV), another member of the family *Bromoviridae*, by
21 cryoelectron microscopy and X-ray crystallography. CMV virions are comprised of 180 copies of
22 a single virus-encoded capsid protein and have T-3 icosahedral symmetry. The 3.2-Å resolution
23 X-ray crystallographic structure of CMV has been described [10, 11]. There are three copies of
24 capsid protein in each icosahedral asymmetric unit of CMV: The A, B and C subunits. This

1 structure shows several differences between the A subunits (taking part in the fivefold axis of
2 symmetry) and the B and C subunits (taking part in the threefold axis of symmetry). The biological
3 assembly of CMV CP is composed of hexameric and pentameric units (Fig.1) [11].

4 Various interactions between viruses and host proteins during infection have been identified.
5 Therefore identifying and describing protein-protein interactions can improve our knowledge of
6 host-virus relationships (Brito and Pinney, 2017; Salánki e al., 2018). In addition, the study of Ma
7 et al. (2008) showed that decreased levels of Ferredoxin I (Fd I) occurred in TMV-infected tobacco



18 **Fig.1.** The schematic structure of the CMV particle. The red and pink represent hexameric
19 structures (B and C subunits) and green represents pentameric structures (A subunits).

20
21 Leaves and may contribute to the expression of chlorosis and mosaic symptoms. Based on previous
22 studies, even the smallest changes in the virus CP can dramatically influence transmission, and
23 also changes in the transmission phenotype can be affected fundamentally, depending on the

1 species of transmitting aphid (Liu et al., 2002; Perry et al., 1998; Zhang et al., 1994). Mutations in
2 the CP often are associated with the evolving nature of symptoms. For instance, the amino acid
3 located in position 129 of the CMV CP has been reported as a critical factor in chlorosis formation
4 [16, 17], aphid transmissibility [15], breaking of host resistance [18], local movement [19], and
5 systemic infection [18]. An interesting type of mutation is the D192K mutation in the CMV CP,
6 which caused necrosis. A similar necrosis has been reported for some single amino-acid
7 substitutions at position 129 (proline) in CP mutants of CMV [16].

8 In a recent study, Qiu and co-workers (2018) demonstrated that the interaction between chloroplast
9 Fd I protein and the CP of M-CMV was critical for the yellow chlorosis symptom development of
10 M-CMV but did not occur with the CP of Q-CMV, and the chlorosis symptoms induced by the
11 silencing of Fd I were the same as those resulting from infection by M-CMV. In addition, chlorosis
12 symptoms on *N. tabacum* plants were significantly induced by the M-CMV and the recombinant
13 virus, Q-CMVMCP (Q-CMV containing the CP of M-CMV), while the Q strain of CMV (Q-
14 CMV) and the reciprocal recombinant virus, M-CMVQCP, did not induce symptoms on tobacco
15 plants, 14 days after inoculation. By contrast, the Fd I abundance decreased following infection by
16 M-CMV, but not by Q-CMV. Collectively, the results of the *in vivo* and *in vitro* experiments,
17 based on yeast two-hybrid analysis and bimolecular fluorescence complementation (BiFC) assays,
18 showed that the chlorosis severity symptoms were directly related to the expression of Fd I, which
19 was down-regulated in plants infected by M-CMV but not Q-CMV [20]. These specific differences
20 in pathogenicity and symptom development between these two strains of CMV have prompted us
21 to investigate and analyze protein-protein interactions between the plant Fd I protein and the CPs
22 of M and Q-CMV. Determining the specificity of the interaction of Fd I with the CPs of a virus
23 strain, inducing either severe chlorosis symptoms (M-CMV) or very mild symptoms (Q-CMV),

1 some highlighted positions, with color differences from yellow indicating the degree of sequence
2 divergence.

3

4 **2.2. Structural modeling and validation**

5 The ProSA Z-Scores were obtained by comparing predicted structures against protein structures
6 of the same size obtained by nuclear magnetic resonance and X-ray crystallography (Fig. S1). The
7 Z-Scores of predicted models were equal to -6.12 (Q-CMV CP) and -5.48 (M-CMV CP) (Table
8 1), which were in the range of Z-score of similar-sized protein structures (Fig. S1).

9 The stereochemical quality of a protein model may be judged by a Ramachandran plot obtained
10 from the PROCHECK program, an algorithm that confirms the protein structure in a plot
11 examining its backbone conformation, by showing the phi and psi angles for each residue of a
12 protein. As shown in Table 1, the predicted model of Q-CMV CP indicates 77.4% acceptable in
13 the regions of most favored and 18.6% in the additional allowed regions, with 2.4% in the
14 generously allowed regions and 1.6% in the disallowed conformations; the predicted model of M-
15 CMV CP indicates 77.7% acceptable in the regions of most favored and 17.9% in the additional
16 allowed regions, 2.4% in the generously allowed regions and 2% in the disallowed conformations.

TABLE 1: Evaluating protein structures modeled by ProSA and PROCHECK (Fig. S1)

	ProSA	PROCHECK regions			
	Z-Score	Most favored %	Additional allowed %	Generously allowed %	Disallowed %
Q-CMV	-6.12	77.4	18.6	2.4	1.6
M-CMV	-5.48	77.7	17.9	2.4	2

17

18 **2.3. MD simulation**

1 The Fny-CMV CP structure was used to model the structures of the pentamer complex (A subunit)
2 and as well as the hexamer complex (B and C subunits) of Q- and M-CMV CP. In order to do
3 further computational studies, the refinement of predicted 3D protein models is important in
4 bringing them closer to experimental accuracy. Hence, all structures of Q- and M-CMV CP,
5 including pentameric and hexameric, were done as MD simulation by GROMAX over 50 ns
6 (Fig.3). Modeled structures of Q- and M-CMV CP, including pentameric and hexameric are shown
7 in Fig 4.

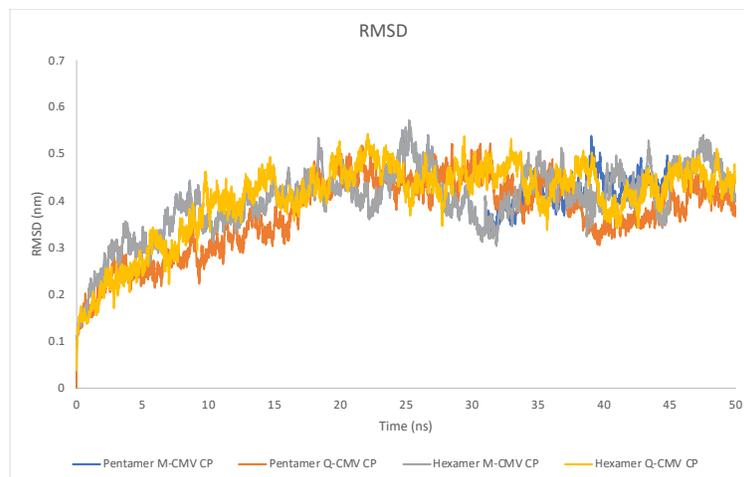
8

9

10

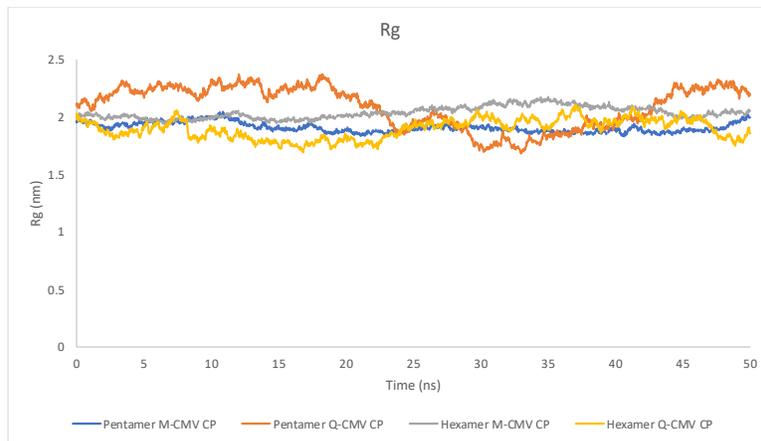
11

12

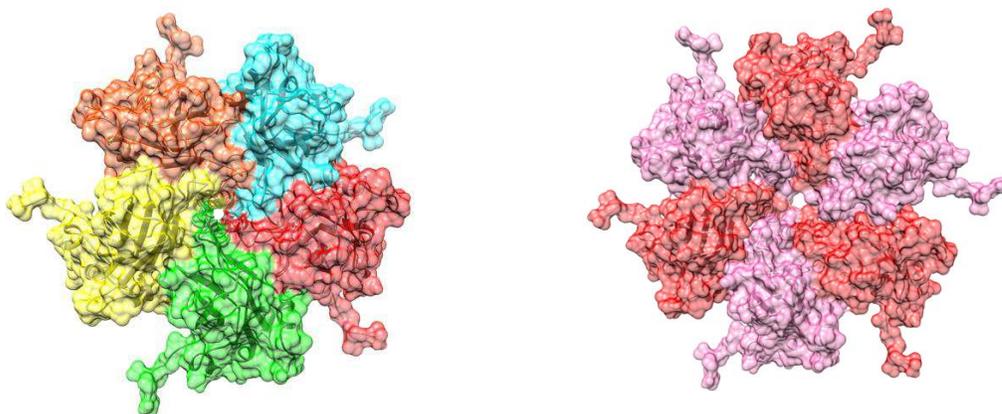


13

14



1
2
3 **Fig. 3.** RMSD analysis of the pentamer and hexamer of Q- and M-CMV CP show that the value
4 of RMSD changes has reached a stationary shape after 20 ns. Also, changes in the Radius of
5 gyration (Rg) shows that there are no significant changes in the size of the simulated protein in the
6 last 30 ns of simulation. These two analyses have shown that the system during the last 30 ns of
7 simulation remains structurally stable and has reached equilibrium (Table S1).

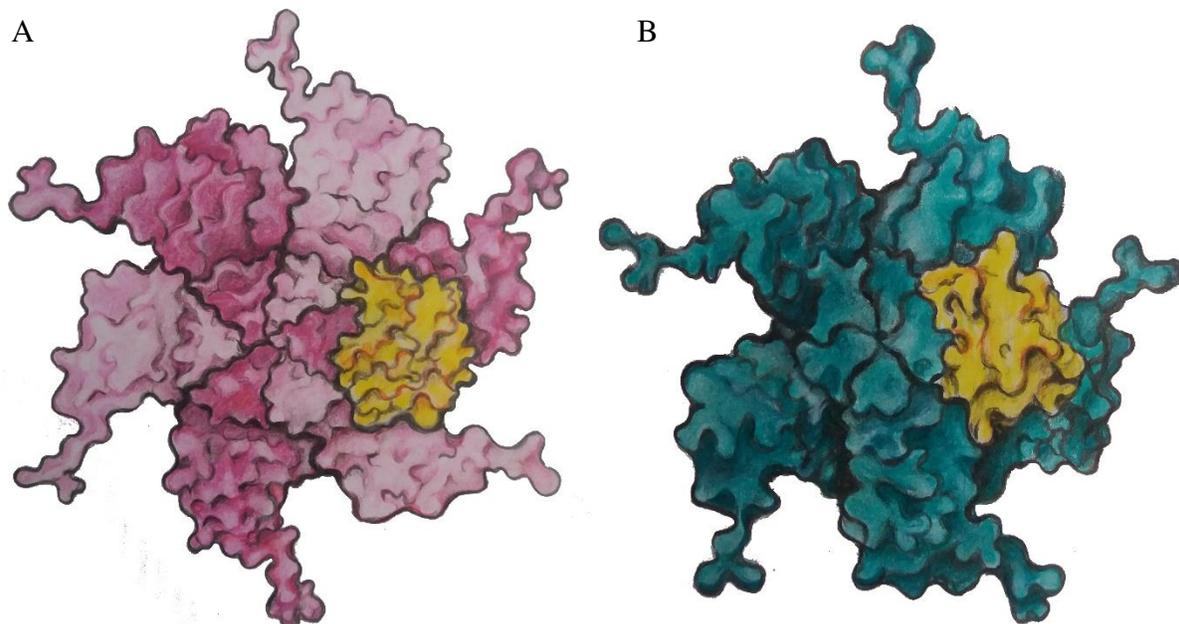


8
9 **Fig.4.** The same structures of Q-CMV and M-CMV CP including pentameric (left) and hexameric
10 (right) forms where each color represents the subunits in the pentameric or hexameric states of the
11 Q-CMV and M-CMV CP.

12

13 **2.4. Interaction of Fd I with Q-CMV or M-CMV CP**

1 As mentioned before, the biological assembly of CMV CPs is available in the PDB database that
2 indicate a single chain is not energetically stable and the quaternary structure for this protein is
3 stable in hexameric and pentameric forms [11]. As is shown schematically in Fig.5, in the
4 estimation of the interactions between the CMV CPs (M and Q strains) with Fd I, the interaction
5 of Fd I with both complex structures (pentameric and hexameric forms of the CP) of the M and Q
6 strains were calculated and no differences between the interactions was detected. The results of
7 the interactions calculated by ClusPro are shown in Table 2. Interestingly, the significant
8 difference between pentameric structures of M-CMV CPs and Fd I compared with pentameric and
9 hexameric forms of Q-CMV CPs and also hexameric structures of M-CMV CPs show that
10 pentameric structures of M-CMV CPs interact with a higher affinity. Despite the significant
11 difference between the interactions of this complex, but due to the high uncertainty of the docking
12 method, affinity was also calculated during the MD simulation.



13
14 **Fig.5.** Mapping of the interaction site between Fd I with (A) hexameric, and (B) pentameric
15 structure of CMV CPs is shown schematically with pink, dark cyan, and yellow colors representing
16 the pentameric CPs, hexameric CPs and Fd I, respectively.

1

TABLE 2: DOCKING RESULTS OF M- AND Q-CMV CPS WITH Fd I BY USING CLUSPRO.

	Pentameric		Hexameric	
	# in cluster	Weighted score	# in cluster	Weighted score
M-CMV CP	194	-1121	69	-765
Q-CMV CP	129	-740	107	-683

2 * The number of members in cluster and weighted score for pentameric and hexameric forms complexed with Fd I are
3 compared between M and Q strains. Both the hexameric form and pentameric form of the Q strain show the same
4 weighted score while the weighted score of the pentameric form of the M strain is significantly higher than the others.

5
6
7 To optimize the complex structures of the docking results of ClusPro, the best clusters were chosen
8 for MD simulation between pentameric and hexameric forms of the M- and Q-CMV CP with Fd
9 I, in which the MD simulations were performed to calculate a more accurate affinity of the
10 complexes for 50 ns (Table S1). The binding affinities obtained by the MMPBSA approach (Table
11 3) have demonstrated the significant differences between binding-free energy of pentameric
12 structures of M-CMV CPs and Fd I compared with pentameric and hexameric forms of Q-CMV
13 CPs and also hexameric structures of M-CMV CPs, therefore, it seems that probably pentameric
14 structures of M-CMV CPs plays a critical role in this interaction.

15

TABLE 3: Binding-free energy (kJ/mol) for the Q-CMV and M- CPs and Fd I conformation by MM-PBSA

Complex	Structure	ΔE_{vdw}	ΔE_{ele}	ΔG_{ps}	ΔG_{SASA}	$\Delta G_{binding}$
M-CMV with Fd I	Pentameric	-331.41+/-0.37	-1476.68+/-0.76	198.11+/-0.53	-82.85+/-0.98	-1610.98+/-0.13
Q-CMV with Fd I		-150.32+/-0.62	-743.45+/-0.06	183.73+/-0.75	-37.58+/-0.09	-709.04+/-0.56
M-CMV with Fd I	Hexameric	-157.6+/-0.32	-693.47+/-0.08	188.32+/-0.33	-39.4+/-0.47	-661.75+/-0.46
Q-CMV with Fd I		-143.77+/-0.54	-712.51+/-0.17	205.87+/-0.45	-35.94+/-0.43	-651.41+/-0.2

Vdw = van der Waals interaction energy; ele = electrostatic interaction energy; ps = polar solvation energy; SASA = nonpolar solvation energy.

3. Discussion:

1 The interactions between the virus and host plant determine the disease development. On the viral
2 side, qualitative and quantitative pathogenicity factors are required in the development of the
3 disease. Plant viruses encode at least the essential proteins for infecting plants and the majority of
4 plant viral proteins work as pathogenicity factors [21]. It has been shown for CMV that satellite
5 RNA and the proteins encoded by all of the genes (i.e., 1a, 2a, 2b, 3a, and CP) act as virulence
6 factors [22]. Previous studies with different CMV strains have shown that CP causes a range of
7 symptom responses including necrosis, mosaics and chloroses, such that substitutions in the amino
8 acid sequence at particular residues changes the severity of the symptoms [23]. The replicase of
9 turnip yellow mosaic virus (a member of the family *Potyviridae*) and viral CP and viral RNA of
10 turnip mosaic virus (a member of the family *Tymoviridae*) are regularly associated with the
11 chloroplast membrane during infection [24, 25]. Both *in vitro* and *in vivo*, during viral infection of
12 potato virus X (PVX) in *Nicotiana* spp., PVX CP-interacting protein 1 (NbPCIP1) from *N.*
13 *benthamiana*, together with the PVX CP, is a key factor for the development of the infection
14 process [26]. Different studies have reported accumulation levels of TMV CP in chloroplasts of
15 infected plants and the severity of chlorosis/mosaic leaves was correlated with the accumulated
16 level of CP in the chloroplasts [27, 28]. More detailed studies have revealed that in light-green
17 areas of TMV-infected plants the amount of CP accumulated in the chloroplast is much more than
18 that in the chloroplast of dark-green regions.

1 Thus, the results of this study showed the nature of a critical interaction between the CP of CMV
2 with the chloroplast Fd I protein. Obviously, MD simulation analyses, particularly MMPBSA,
3 have shown pentameric structures of M-CMV CPs strongly interact with Fd I, while such an
4 interaction was not observed for hexameric structures of M-CMV CPs and also pentameric and
5 hexameric forms of Q-CMV CPs. The more detailed MMPBSA has shown that the electrostatic
6 bonds play a key role in pentameric complex form of M-CMV CPs with Fd I. It appears that the
7 differences in symptom development associated with the CPs of M- and Q-CMV are related to
8 how Fd I is associated with the CPs of these two strains. It is likely that the activity of plant Fd I
9 is inhibited by trapping it in the complex of pentameric structures and hence, Fd I will not be able
10 to function in the photosynthetic electron transport chain. Evaluating the content of CP and Fd I
11 by western blot analysis revealed that the amount of CP increased as the chlorosis symptoms
12 became more severe, whereas the abundance of Fd I diminished during M-CMV infection (Qiu et
13 al. 2018). Meanwhile, it was demonstrated that the amount of Fd I in Q-CMV–infected samples
14 was not diminished, compared to the level of Fd I in mock-inoculated samples. On the other hand,
15 the northern blot analysis of Fd I mRNA content did not show any differences between all tested
16 cases (Qiu et al. 2018). Hence, Fd I mRNA accumulation was not affected by viral infection. Thus,
17 the decrease in the abundance of Fd I may be related to interaction with the CP of M-CMV.
18 Silencing Fd I expression in *N. benthamiana* plants, by using the tobacco rattle virus based virus-
19 induced gene silencing system expressing a partial clone of the *Fd I* gene, produced chlorosis
20 symptoms, similar to those induced by M-CMV infection (Qiu et al. 2018). Thus, the results of the
21 current study were consistent with the study of Qiu et al. (2018) confirming the critical interaction
22 between the CP of M-CMV with chloroplast Fd I proteins. Interestingly, Bowman et al. (2002)
23 demonstrated that the pentameric structure of the A subunits, as a recognition site, was involved

1 in aphid transmission of CMV, whereas such roles for the hexameric structures of the B and C
2 subunits were not detected. Similarly, the results of our study showed a strong interaction between
3 pentameric structures of the CP of M-CMV with Fd I in comparison with those of Q-CMV. The
4 affinity of the CPs of the M and Q strains for Fd I is approximately the same in the hexamer state
5 and is about the same as the affinity of Fd I for the pentamer of the CP of Q-CMV. Trapping Fd I
6 in the complex of pentameric structures can lead to the inhibition of the activity of plant Fd I in
7 the photosynthetic electron transport chain. Consequently, as opposed to the situation with Q-
8 CMV and Fd I, symptom development is induced by M-CMV.

9 The findings of Zhang et al. (1994) indicate that the severity of systemic symptoms in tobacco is
10 determined by RNAs 1 and 2 of Fny-CMV. However, the study by Lewsey et al. (2009) showed
11 that the 2b protein was an important symptom determinant in certain hosts, while the study of
12 González et al. (2010) indicated that the 2b protein worked synergistically with other CMV
13 components to produce viral symptoms. Surprisingly, *Arabidopsis thaliana* ecotype C24 plants
14 infected with Fny-CMV Δ 2b showed severe symptoms similar to those seen with wild-type virus
15 infection. Although extensive empirical studies have demonstrated that the M-CMV and TMV
16 CPs have also played a critical role in viral symptom development [20, 28], the existence of these
17 differences helps us to avoid simplistic assumptions that symptoms of CMV infection should be
18 attributed to a single gene product. As a result, the suggestion of any part of the CMV genome
19 alone such as 2b, CP, etc., as the sole cause of CMV symptoms would be incorrect. Hence, it would
20 be more logical to presume that the symptoms induced by CMV can be different, depending on
21 the viral proteins responsible for inducing symptom in infecting virus strains and the host being
22 infected.

1 Many previous studies have shown that different types of biotic and abiotic stresses change the
2 level of Fd I. In other hands, several studies [29–33] have indicated that transgenic plants with
3 increasing levels of Fd I have shown enhanced resistance against bacterial pathogens [29, 30].
4 These studies have shown that Fd I level is correlated to the resistance of plants to stress conditions.
5 Accordingly, Fd I alteration in the level of expression may modify the resistance of plants to
6 stresses (Ger et al., 2002; Tang et al., 2001b; Tognetti et al., 2006). Decreasing the level of Fd I
7 by antisense Fd I-1 or virus-induced RNA silencing of the *Fd I-1* gene in *N. benthamiana* has
8 indicated that the level of CP in the chloroplasts of TMV–infected transgenic plants remained
9 considerably higher than that in the chloroplasts of wild type control plants. But the results of the
10 more comprehensive study in the light-green areas with higher TMV levels and dark-green areas
11 with lower TMV levels have shown that the Fd I level in the light-green regions has remained
12 significantly lower than in the dark-green regions. Nevertheless, interestingly, the level of Fd I in
13 the light and dark-green regions were both lower than in the healthy leaves. Ma et al. (2008) has
14 reported that in fact the reduction of the level of Fd I in TMV-infected tobacco plants was indeed
15 caused by the infection of TMV. We suggest that the symptom development model presented for
16 CMV can help us to describe more precisely the symptom development model in TMV. First, it
17 is suggested that like what was seen for CMV infection, probably occurs in TMV–infected plants
18 with CPs interacting with Fd I. Second, in the light-green areas with higher TMV-infection, Fd I
19 activity may be inhibited by CP and subsequently antibody against Fd I cannot detect it well.
20 However, in the dark-green areas containing much less TMV-infection, Fd I is probably less
21 inhibited by CP and subsequently antibody against Fd I can detect it very well.

22

23 4. Conclusions:

1 In this study, the *in silico* interaction between pentameric and hexameric structures of CP of CMV
2 and FdI protein has been shown. A further study comparing the two strains [severe chlorosis
3 symptoms (M-CMV) and mild symptoms (Q-CMV)] along with more details in order to
4 understand these interactions have shown that in M-CMV the pentameric structures of CP strongly
5 interact with plant chloroplast Fd I protein. By contrast, a strain with lower pathogenicity and
6 milder symptoms, Q strain, was not found to have such as relationship for either the pentamer or
7 hexamer structures. Therefore, our study suggests a model for symptom development in CMV,
8 based on the differences in interaction of Fd I with pentameric structures of CMV CPs, which can
9 lead to inhibition of the activity of plant Fd I in the photosynthetic electron transport chain, with
10 symptoms development is seen more severe in plants infected by M-CMV than Q-CMV. Based
11 on this model any factor that inhibits the interaction of CMV CPs with plant Fd I can prevent the
12 symptom development and eventually resistance will emerge.

13 **5. Methods:**

14 **2.1. Multiple Sequence Alignment and Functional Annotation:**

15 The protein sequences of Fd I (<https://www.uniprot.org/uniprot/Q6Q8B8>) and the CPs of Fny-
16 CMV (<https://www.uniprot.org/uniprot/P69466>), M-CMV (<https://www.uniprot.org/uniprot/Q00260>),
17 and Q-CMV CP (<https://www.uniprot.org/uniprot/P03605>), were obtained from UniProt. The
18 online version of Clustal Omega (with default parameters) was used for aligning CP sequences
19 (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). The Jalview software was used to display multiple
20 sequence alignments [36]. The available 3-D structure of Fny-CMV CP and Fd I, determined by
21 X-ray crystallography, were obtained from Protein Data Bank (PDB: <https://www.rcsb.org/>)
22 entries 1F15 and 4ZHP, respectively.

2.2. Homology modeling and model optimization

1 Given that the structural information for the Q-CMV and M-CMV CPs were not available in the
2 PDB database, the Fny-CMV CP (1f15) was selected as a template because of its high sequence
3 similarity to these target proteins. Hence, to have a three-dimensional structure of Q-CMV and M-
4 CMV strains, homology modeling was performed using the modeling package MODELLER ver.
5 9.22. Based on their PDB entries, the three-dimensional structure of Fny-CMV is a homotrimer
6 structure that includes the 218 amino-acid capsid protein, forming A, B, and C subunits. The
7 structural information of the N-terminal 28 amino acid residues suggests that these are mostly
8 disordered for all subunits, while the structural information for residues 28-62 of the A subunit is
9 not available either. It is, therefore, possible to complete the structural information of subunit A
10 based on structural information of either subunits B or C. Hence, the residues from 28 to 62 of the
11 A subunit were modeled by MODELLER software [37], based on equivalent residues of the B
12 subunit. The homology modeling of the N-terminal residues 28 to 62 of a subunit will provide a
13 complete structure of the viral protein.

14 The quality of the predicted structure was analyzed using structural validation algorithms ProSA.
15 To further verify the best-achieved 3D model, the structure was evaluated by a Ramachandran plot
16 within the PROCHECK server [38]. The Z-score of the ProSA webserver was used for measuring
17 the sequence-structure compatibility [39]. Optimization of the model energy was achieved using
18 energy minimization protocols available at GROMACS [40] software.

19 **2.3. Molecular docking**

20 In order to facilitate understanding the nature of the interaction between the CPs of either M-CMV
21 or Q-CMV with Fd I, the molecular docking was done using the ClusPro 2.06 program [41]. After
22 docking, the models were built by ClusPro 2.06, using default parameters. The weighted scores
23 were calculated based on $E=0.40E_{rep}+0.40E_{att}+600E_{elec}+1.00E_{DARS}$ for all models and their

1 related clusters; afterwards, the best-offered cluster with the most members and the lowest
2 weighted scores docking mode, was selected (cluster #0). The predicted models of ClusPro are
3 ranked by cluster size. Finally, the best models with the lowest weighted scores were selected and
4 analyzed further. The molecular docking procedure allows a deeper insight into the interaction
5 between Fd I and the CP of either Q-CMV or M-CMV CP at the molecular level.

6

7 **2.4. Molecular dynamics (MD) simulation and binding free energy calculation**

8 In this study, the MD techniques were applied by using the charmm36 force field [42] to appraise
9 the stability and consistency of the obtained models. Some independent simulations were done by
10 GROMACS, for the modeled pentameric structures, including A subunits, and hexameric
11 structures, including B and C subunits, for both M and Q-CMV CPs. To perform MD simulations,
12 the most likely model generated by ClusPro between complex M and Q strains of CMV CPs,
13 including pentameric and hexameric forms with Fd I were selected. Each complex structure was
14 entered in a cubic box in which at least a minimum 1 nm distance was chosen between a cubic box
15 of molecules and the proteins, (11.38 nm, 11.38 nm, 11.38 nm), which contains the CP solvated
16 with 38442 SPC216 water molecules and neutralized by adding 0.15 M NaCl. To simulate the
17 whole system by replicating the initial cell, the three-dimensional periodic boundary conditions
18 (PBCs) were regarded to dispose of boundary effect complications. The system was subjected to
19 energy minimization by using the steepest descent algorithm ($F_{max} < 1000$). All complexes were
20 simulated to equilibrate at 300°K temperature using the v-rescale method [43] and were kept at 1
21 bar, based on the Berendsen method [44]. The systems were submitted into a 50 ns (single
22 structures of pentameric and hexameric) and 50 ns (complex structures of pentameric and

1 hexameric with Fd I) MD simulations equilibrated in the canonical (NVT) and isothermal–isobaric
 2 (NPT) ensembles, using the leapfrog algorithm with an integration time step of 0.002 ps.
 3 Finally, the Molecular Mechanics Poisson Boltzmann Surface Area (MM-PBSA) approach for
 4 calculation free energy was applied the g-mmpbsa script [45], during the last 30 ns period of MD
 5 trajectory. It was calculated as the aggregate of (Electrostatic + Van der Waal), Solvent Accessible
 6 Surface Area (SASA), and (polar + non-polar solvation energy). Seven parameters were used for
 7 MM-PBSA calculations. The formula can be defined by the calculations of the binding free energy,
 8 based on the following equations:

$$9 \quad \Delta G_{\text{bind}} = G_{\text{comp}} - (G_{\text{recep}} + G_{\text{lig}})$$

10 Therefore G_{comp} , G_{recep} and G_{lig} are equal to binding free energy for complex, unbound receptor,
 11 and unbound ligand respectively. The free binding energy for each of the equation calculation
 12 was determined as given below:

$$13 \quad G_x = E_{\text{bond}} + (E_{\text{elec}} + E_{\text{vdW}}) + G_{\text{polar}} + \gamma \text{SASA} + b$$

14 Where G_x is calculated as (G_{comp} , G_{recep} or G_{lig}). E_{bond} always consider zero, and $E_{\text{elec}} + E_{\text{vdW}}$
 15 contain electrostatic and van der Waal interactions. G_{polar} and $G_{\text{non-polar}}$ are as the electrostatic free
 16 solvation energy. γ is the co-efficient of Solvent-accessible surface area (SASA), and b as a
 17 fitting parameter [46].

18 **List of abbreviations**

19 **CMV:** Cucumber mosaic virus

20 **Q-CMV:** Q strain of CMV

21 **M-CMV:** M strain of CMV

22 **Fd I:** Ferredoxin I

23 **NbPCIP1:** *Nicotiana* spp., PVX CP-interacting protein 1

1 **MD:** Molecular dynamics

2 **PDB:** Protein Data Bank

3 **MMPBSA:** Molecular Mechanics Poisson-Boltzmann Surface Area

4 **V_{dw}** : van der Waals interaction energy

5 **ele:** electrostatic interaction energy

6 **ps:** polar solvation energy

7 **SASA:** Solvent Accessible Surface Area

8

9

10

11 **Declarations:**

12 **Ethical approval**

13 This article does not contain any studies involving animals performed by any of the authors.

14 **Consent for publication**

15 Not applicable

16 **Availability of data and materials**

17 All data generated or analyzed during this study, including raw sequence files, are included in

18 this article.

19 **Declaration of competing interest**

20 The authors have declared no conflicts of interest

21 **Funding**

22 This research did not receive any specific grant from funding agencies in the public, commercial,

23 or not-for-profit sectors.

1 **Authors' contributions**

2 Akbari motlagh M., Shams-Bakhsh M., Arab S. S., conceived designed, and performed the
3 computational test; Akbari motlagh M., Arab S. S. analyzed the data; Akbari motlagh M. wrote
4 the main manuscript text; Shams-Bakhsh M., Arab S. S., Palukaitis P. revised the manuscript.

5 **Acknowledgements**

6 The authors would like to acknowledge the Zahra Tahershamsi very much for illustration of the
7 schematic structure of the CMV particle (Fig 1 and 5). We also thank Tarbiat Modares
8 University for providing the necessary facilities to carry out the bioinformatics analysis.

9

10 **References:**

- 11 1. Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. Virus
12 taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV).
13 Nucleic Acids Res. 2018;46:D708–17. doi:10.1093/nar/gkx932.
- 14 2. Scholthof KBG, Adkins S, Czosnek H, Palukaitis P, Jacquot E, Hohn T, et al. Top 10 plant
15 viruses in molecular plant pathology. Mol Plant Pathol. 2011;12:938–54.
- 16 3. Qiao Y, Li HF, Wong SM, Fan ZF. Plastocyanin transit peptide interacts with *Potato virus X*
17 coat protein, while silencing of plastocyanin reduces coat protein accumulation in chloroplasts
18 and symptom severity in host plants. Mol Plant-Microbe Interact. 2009;22:1523–34.
19 doi:10.1094/MPMI-22-12-1523.
- 20 4. Jacquemond M. Cucumber mosaic virus. In: Advances in Virus Research. Academic Press
21 Inc.; 2012. p. 439–504. doi:10.1016/B978-0-12-394314-9.00013-0.
- 22 5. Yoon J-Y, Palukaitis P, Choi S-K. CHAPTER 1: Host range. in: cucumber mosaic virus. The
23 American Phytopathological Society; 2019. p. 15–8. doi:10.1094/9780890546109.004.

- 1 6. Palukaitis P, García-Arenal F. Cucumoviruses. *Adv Virus Res.* 2003;62:241–323.
- 2 <http://www.ncbi.nlm.nih.gov/pubmed/14719367>. Accessed 5 Oct 2019.
- 3 7. Stevens WA. *Virology of flowering plants.* Springer US; 1983.
- 4 8. Lewsey MG, González I, Kalinina NO, Palukaitis P, Canto T, Carr JP. Symptom induction
- 5 and RNA silencing suppression by the cucumber mosaic virus 2b protein. *Plant Signal Behav.*
- 6 2010;5:705–8.
- 7 9. Krenz B, Bronikowski A, Lu X, Ziebell H, Thompson JR, Perry KL. Visual monitoring of
- 8 cucumber mosaic virus infection in *nicotiana benthamiana* following transmission by the aphid
- 9 vector *myzus persicae*. *J Gen Virol.* 2015;96:2904–12. doi:10.1099/vir.0.000185.
- 10 10. Speir JA, Munshi S, Wang G, Baker TS, Johnson JE. Structures of the native and swollen
- 11 forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron
- 12 microscopy. *Structure.* 1995;3:63–78. doi:10.1016/s0969-2126(01)00135-6.
- 13 11. Smith TJ, Chase E, Schmidt T, Perry KL. The structure of cucumber mosaic virus and
- 14 comparison to cowpea chlorotic mottle virus. *J Virol.* 2000;74:7578–86.
- 15 doi:10.1128/jvi.74.16.7578-7586.2000.
- 16 12. Brito AF, Pinney JW. Protein-protein interactions in virus-host systems. *Front Microbiol.*
- 17 2017;8 AUG:1–11.
- 18 13. Liu S, He X, Park G, Josefsson C, Perry KL. A conserved capsid protein surface domain of
- 19 cucumber mosaic virus is essential for efficient aphid vector transmission. *J Virol.*
- 20 2002;76:9756–62.
- 21 14. Zhang L, Hanada K, Palukaitis P. Mapping local and systemic symptom determinants of
- 22 cucumber mosaic cucumovirus in tobacco. *J Gen Virol.* 1994;75:3185–91.
- 23 15. Perry KL, Zhang L, Palukaitis P. Amino acid changes in the coat protein of cucumber mosaic

- 1 virus differentially affect transmission by the aphids *Myzus persicae* and *Aphis gossypii*.
2 *Virology*. 1998;242:204–10. doi:10.1006/viro.1998.8991.
- 3 16. Suzuki M, Kuwata S, Masuta C, Takanami Y. Point mutations in the coat protein of
4 cucumber mosaic virus affect symptom expression and virion accumulation in tobacco. *J Gen
5 Virol*. 1995;76:1791–9. doi:10.1099/0022-1317-76-7-1791.
- 6 17. Shintaku MH, Zhang L, Palukaitis P. A single amino acid substitution in the coat protein of
7 cucumber mosaic virus induces chlorosis in tobacco. *Plant Cell*. 1992;4:751–7.
8 doi:10.1105/tpc.4.7.751.
- 9 18. Ryu KH, Kim C-H, Palukaitis P. The coat protein of cucumber mosaic virus is a host range
10 determinant for infection of maize. *Mol Plant-Microbe Interact*. 1998;11:351–7.
11 doi:10.1094/MPMI.1998.11.5.351.
- 12 19. Wong S-M, Thio SS-C, Shintaku MH, Palukaitis P. The rate of cell-to-cell movement in
13 squash of cucumber mosaic virus is affected by sequences of the capsid protein. *Mol Plant-
14 Microbe Interact*. 1999;12:628–32. doi:10.1094/MPMI.1999.12.7.628.
- 15 20. Qiu Y, Zhang Y, Wang C, Lei R, Wu Y, Li X, et al. Cucumber mosaic virus coat protein
16 induces the development of chlorotic symptoms through interacting with the chloroplast
17 ferredoxin I protein. *Sci Rep*. 2018;8:1205. doi:10.1038/s41598-018-19525-5.
- 18 21. Hull R. *Plant Virology* (5th ed.) by Hull, Roger . Academic Press . 2013.
19 <https://www.ebooks.com/en-us/book/1562324/plant-virology/roger-hull/>. Accessed 28 Aug
20 2020.
- 21 22. Mochizuki T, Ohki ST. Single amino acid substitutions at residue 129 in the coat protein of
22 cucumber mosaic virus affect symptom expression and thylakoid structure. 2011;:881–6.
- 23 23. Mochizuki T, Ohki ST. Cucumber mosaic virus: Viral genes as virulence determinants. *Mol*

- 1 Plant Pathol. 2012;13:217–25.
- 2 24. Wei T, Zhang C, Hong J, Xiong R, Kasschau KD, Zhou X, et al. Formation of complexes at
3 plasmodesmata for potyvirus intercellular movement is mediated by the viral protein P3N-PIPO.
4 PLoS Pathog. 2010;6:e1000962. doi:10.1371/journal.ppat.1000962.
- 5 25. Prod'homme D, Jakubiec A, Tournier V, Drugeon G, Jupin I. Targeting of the turnip yellow
6 mosaic virus 66K replication protein to the chloroplast envelope is mediated by the 140K
7 protein. J Virol. 2003;77:9124–35. doi:10.1128/jvi.77.17.9124-9135.2003.
- 8 26. Park M-R, Park S-H, Cho S-Y, Kim K-H. *Nicotiana benthamiana* protein, NbPCIP1,
9 interacting with Potato virus X coat protein plays a role as susceptible factor for viral infection.
10 Virology. 2009;386:257–69. doi:10.1016/J.VIROL.2008.12.044.
- 11 27. Banerjee N, Wang J-Y, Zaitlin M. A single nucleotide change in the coat protein gene of
12 tobacco mosaic virus is involved in the induction of severe chlorosis. Virology. 1995;207:234–9.
13 doi:10.1006/viro.1995.1070.
- 14 28. Ma Y, Zhou T, Hong Y, Fan Z, Li H. Decreased level of ferredoxin I in Tobacco mosaic
15 virus-infected tobacco is associated with development of the mosaic symptom. Physiol Mol Plant
16 Pathol. 2008;72:39–45. doi:10.1016/J.PMPP.2008.05.004.
- 17 29. Huang HE, Ger MJ, Chen CY, Pandey AK, Yip MK, Chou HW, et al. Disease resistance to
18 bacterial pathogens affected by the amount of ferredoxin-I protein in plants. Mol Plant Pathol.
19 2007;8:129–37.
- 20 30. Tang K, Sun X, Hu Q, Wu A, Lin C-H, Lin H-J, et al. Transgenic rice plants expressing the
21 ferredoxin-like protein (AP1) from sweet pepper show enhanced resistance to *Xanthomonas*
22 *oryzae* pv. *oryzae*. Plant Sci. 2001;160:1035–42. doi:10.1016/S0168-9452(01)00351-X.
- 23 31. Kim M, Lee S, Park K, Jeong EJ, Ryu CM, Choi D, et al. Comparative microarray analysis

- 1 of programmed cell death induced by proteasome malfunction and hypersensitive response in
2 plants. *Biochem Biophys Res Commun.* 2006;342:514–21.
- 3 32. Zou J, Rodriguez-Zas S, Aldea M, Li M, Zhu J, Gonzalez DO, et al. Expression profiling
4 soybean response to *Pseudomonas syringae* reveals new defense-related genes and rapid HR-
5 specific downregulation of photosynthesis. *Mol Plant-Microbe Interact.* 2005;18:1161–74.
- 6 33. Tognetti VB, Palatnik JF, Fillat MF, Melzer M, Hajirezaei MR, Valle EM, et al. Functional
7 replacement of ferredoxin by a cyanobacterial flavodoxin in tobacco confers broad-range stress
8 tolerance. *Plant Cell.* 2006;18:2035–50.
- 9 34. Ger, Mang Jye Chen, Cheng Hsien Hwang, Shaw Yhi Huang, Hsiang En Podile, Appa Rao
10 Dayakar, Badri Venkata Feng TY. Constitutive expression of hrap gene in transgenic tobacco
11 plant enhances resistance against virulent bacterial pathogens by induction of a hypersensitive
12 response. *Mol Plant-Microbe Interact.* 2002;15:764–73.
- 13 35. Tang K, Sun X, Hu Q, Wu A, Lin C-H, Lin H-J, et al. Transgenic rice plants expressing the
14 ferredoxin-like protein (AP1) from sweet pepper show enhanced resistance to *Xanthomonas*
15 *oryzae* pv. *oryzae*. *Plant Sci.* 2001;160:1035–42. doi:10.1016/S0168-9452(01)00351-X.
- 16 36. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A
17 multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009;25:1189–91.
- 18 37. Šali A. Modelling mutations and homologous proteins. *Curr Opin Biotechnol.* 1995;6:437–
19 51. doi:10.1016/0958-1669(95)80074-3.
- 20 38. Laskowski RA, MacArthur MW, Thornton JM. PROCHECK : validation of protein-structure
21 coordinates . American Cancer Society; 2012. p. 684–7. doi:10.1107/97809553602060000882.
- 22 39. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in
23 three-dimensional structures of proteins. *Nucleic Acids Res.* 2007;35 Web Server:W407–10.

- 1 doi:10.1093/nar/gkm290.
- 2 40. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. Gromacs: High
3 performance molecular simulations through multi-level parallelism from laptops to
4 supercomputers. *SoftwareX*. 2015;1–2:19–25.
- 5 41. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server
6 for protein–protein docking. *Nat Protoc*. 2017;12:255–78. doi:10.1038/nprot.2016.169.
- 7 42. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an
8 improved force field for folded and intrinsically disordered proteins. *Nat Methods*. 2017;14:71–
9 3. doi:10.1038/nmeth.4067.
- 10 43. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem*
11 *Phys*. 2007;126:014101. doi:10.1063/1.2408420.
- 12 44. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics
13 with coupling to an external bath. *J Chem Phys*. 1984;81:3684–90. doi:10.1063/1.448118.
- 14 45. Kumari R, Kumar R, Lynn A. G-mmpbsa -A GROMACS tool for high-throughput MM-
15 PBSA calculations. *J Chem Inf Model*. 2014;54:1951–62. doi:10.1021/ci500020m.
- 16 46. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Kollman, P. A., Massova,
17 I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O.,
18 Cieplak, P., Srinivasan, J., Case, D. A., & Cheatham, T. E. (2000). Calculating structures and
19 free energies of complex molecules: *Comb. Acc Chem Res*. 2000;33:889–97.
20 doi:10.1021/ar000033j.

21

Figures

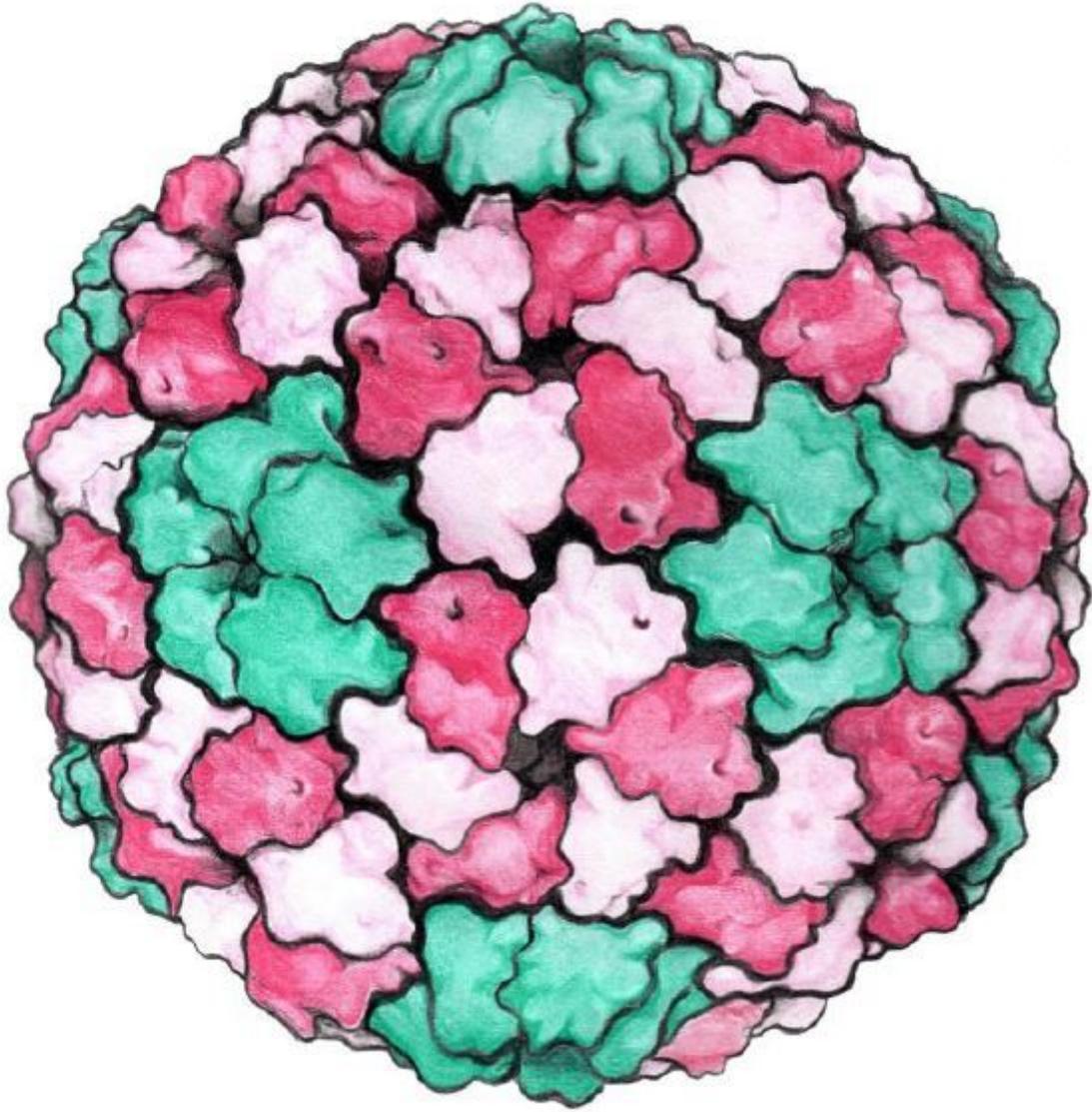


Figure 1

The schematic structure of the CMV particle. The red and pink represent hexameric structures (B and C subunits) and green represents pentameric structures (A subunits).

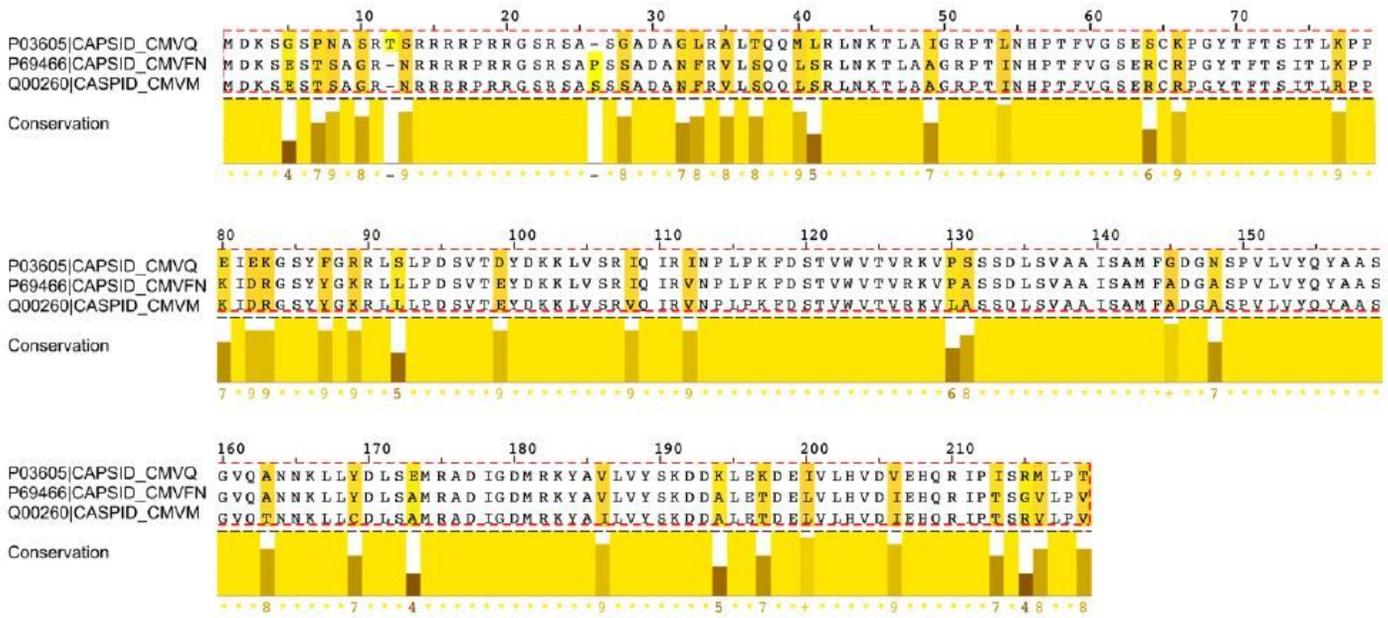


Figure 2

Multiple sequence alignment between CP of Fny strain of CMV and the CPs of the strains M-CMV and Q-CMV. Most of the sequences of these three proteins are identical but differ in some highlighted positions, with color differences from yellow indicating the degree of sequence divergence.

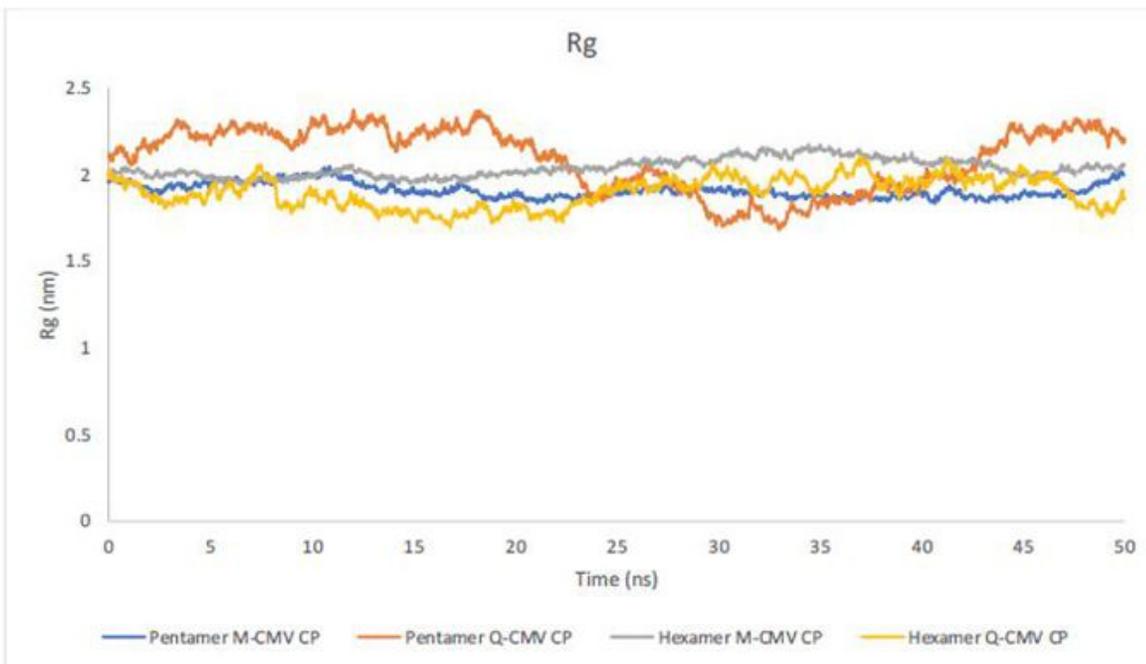
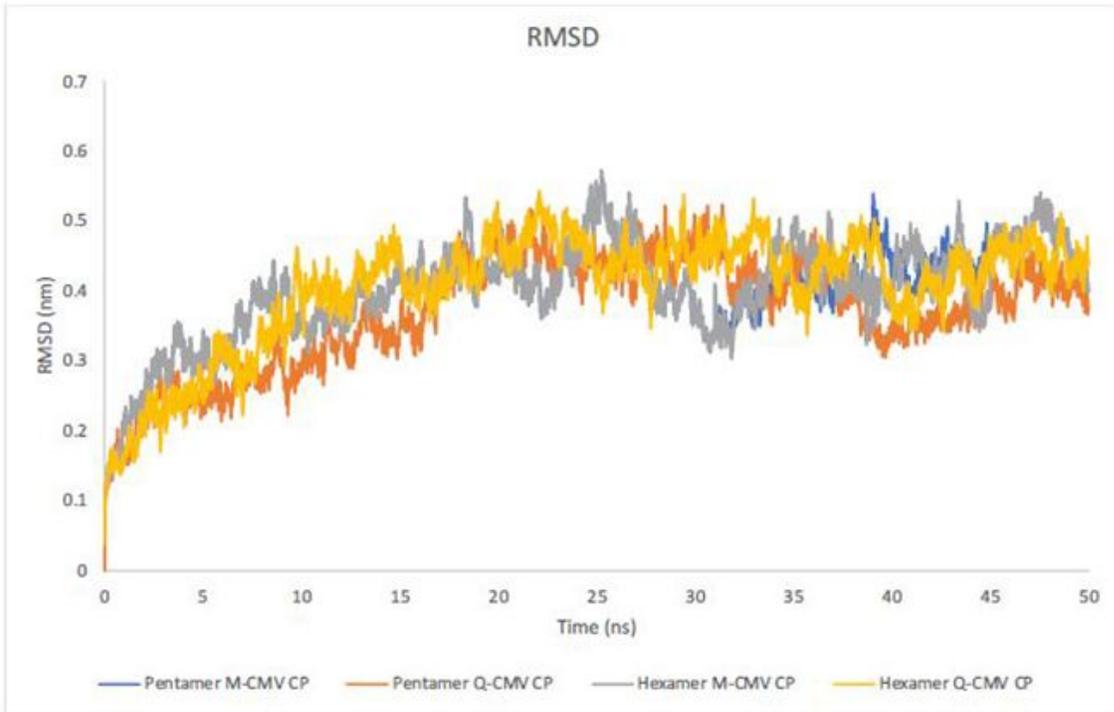


Figure 3

RMSD analysis of the pentamer and hexamer of Q- and M-CMV CP show that the value of RMSD changes has reached a stationary shape after 20 ns. Also, changes in the Radius of gyration (Rg) shows that there are no significant changes in the size of the simulated protein in the last 30 ns of simulation. These two analyses have shown that the system during the last 30 ns of simulation remains structurally stable and has reached equilibrium (Table S1).

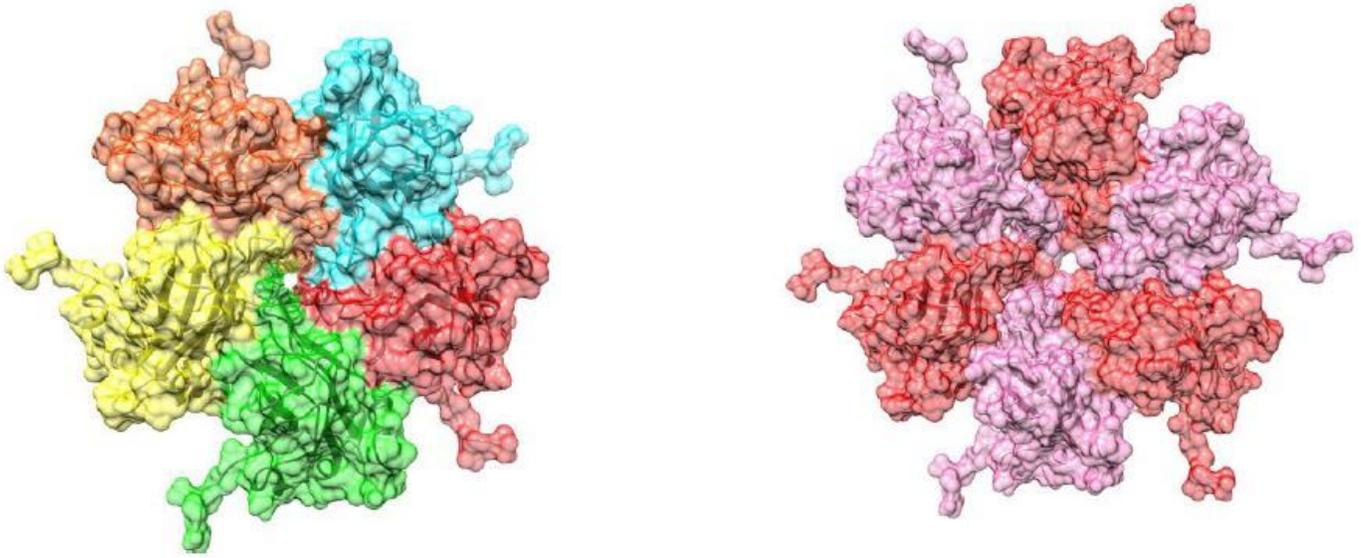


Figure 4

The same structures of Q-CMV and M-CMV CP including pentameric (left) and hexameric (right) forms where each color represents the subunits in the pentameric or hexameric states of the Q-CMV and M-CMV CP.

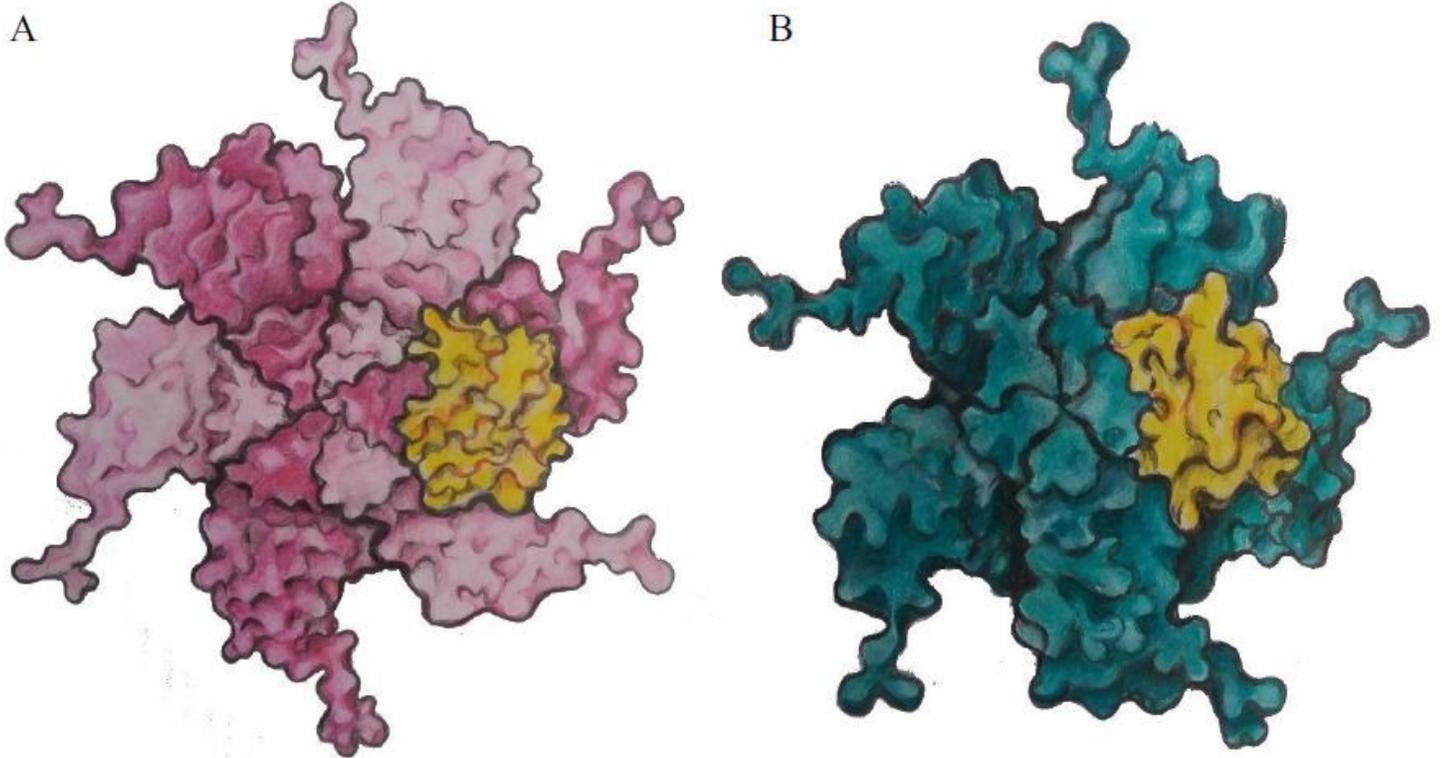


Figure 5

Mapping of the interaction site between Fd I with (A) hexameric, and (B) pentameric structure of CMV CPs is shown schematically with pink, dark cyan, and yellow colors representing the pentameric CPs, hexameric CPs and Fd I, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile2TableS1MDAnalysis.xlsx](#)
- [Additionalfile1figS1.pdf](#)