Journal of Molecular Graphics and Modelling 28 (2010) 569–576

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

# The solution structures of the Cucumber mosaic virus and Tomato aspermy virus coat proteins explored with molecular dynamics simulations

Ákos Gellért *, Ervin Balázs

*Agricultural Research Institute of the Hungarian Academy of Sciences, Department of Applied Genomics, H-2462 Martonvásár, Brunszvik u. 2. Hungary*

ARTICLE INFO

*Article history:*
Received 5 October 2009
Received in revised form 1 December 2009
Accepted 8 December 2009
Available online 21 December 2009

*Keywords:*
Cucumoviruses
*Ab initio* protein modeling
Molecular dynamics simulation
Intrinsically unstructured proteins
Principal component analysis

ABSTRACT

The three-dimensional structures of two *cucumovirus* coat proteins (CP), namely *Cucumber mosaic virus* (CMV) and *Tomato aspermy virus* (TAV), were explored by molecular dynamics (MD) simulations. The N-terminal domain and the C-terminal tail of the CPs proved to be intrinsically unstructured protein regions in aqueous solution. The N-terminal α-helix had a partially unrolled conformation. The thermal factor analysis of the CP loop regions demonstrated that the CMV CP had more flexible loop regions than the TAV CP. The principal component analysis (PCA) of the MD trajectories showed that the first three eigenvectors represented the three main conformational motions in the CPs. The first motion components with the highest variance contribution described an opening movement between the hinge and the N-terminal domain of both CPs. The second eigenvector showed a closing motion, while the third eigenvector represented crosswise conformational fluctuations. These new findings, together with previous results, suggest that the hinge region of CPs plays a central role in the recognition and binding of viral RNA.



## 1. Introduction

*Cucumoviruses* cause serious losses to agriculture worldwide, principally in vegetable and ornamental plants. New protection measures can be envisaged against plant viruses if we explore the detailed molecular mechanism of the infection process. The *Cucumber mosaic virus* (CMV) and *Tomato aspermy virus* (TAV) belong to the *Cucumovirus* genus in the family *Bromoviridae* [1]. Their genomes consist of three single-stranded, positive-sense RNA molecules. Elements of the viral replication system, proteins 1a and 2a, are encoded by RNA 1 and 2 [2]. RNA 2 encodes another small protein (2b), which is responsible for RNA silencing and takes part in the regulation of symptom expression, in a shifted and overlapped open reading frame (ORF). RNA 3 encodes the movement protein (3a, MP) and coat protein (3b, CP). The main function of the CP is to encapsulate the viral RNA. The virus particle of *cucumoviruses* is assembled from 180 CP subunits with a T = 3 truncated icosahedral symmetry [3,4].

When plant viruses infect host plants, viruses spread from the first infected cell to the adjacent cells through the plasmodesmata in a process known as cell-to-cell movement. Cell-to-cell movement requires the viral movement proteins, which regulate the transfer of viral RNAs through the plasmodesmata. In some cases, the MP alone is competent for viral cell-to-cell movement and for virus transfer as a ribonucleoprotein complex (e.g. *Tobacco mosaic virus*, TMV). In other cases, the coat protein is also essential for this process; tubular structures are formed that cut across the cell wall through modified plasmodesmata (e.g. *tospoviruses, comoviruses, potexviruses*), and the virus moves as a virus particle [5,6]. *Cucumoviruses* also require the presence of the monomeric CP for cell-to-cell movement, but not the virus particles [7].

On the basis of previous results [8] and data from the literature, it can be confirmed that the coat proteins of *cucumoviruses* not only build up the virus particle but also play other essential roles in the molecular mechanism of the virus infection processes. Both the movement protein and the coat protein are required for the cell-to-cell movement of *cucumoviruses*. The exchangeability of the R-CMV and P-TAV CP and MP and their hybrid variants was explored [8]. The chimeric virus constructs, pT3NSR and pR3SPT, containing hybrid CP (amino acid (aa) 1–69 CMV CP and aa 70–218 TAV CP), had a non-infectious nature. After the introduction of two point mutations (E62K/K65R) the surface electrostatic potential pattern of the CP was corrected and the mutated chimeric virus construct became infectious. It was assumed that there could be a new functional site on the region surrounding amino acids 62 and 65 of the coat protein surface, which is responsible for an essential step in cell-to-cell movement. These residues are located in the so-called hinge region of the CP. The three-dimensional structures of the CMV and TAV coat protein were explored by a molecular dynamics (MD) simulation. The main aim of this study was to

* Corresponding author. Tel.: +36 22 569 500x317; fax: +36 22 569 514.
*E-mail address:* gellerta@mail.mgki.hu (Á. Gellért).

1093-3263/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2009.12.002

obtain more information on systems in which the initial state is the X-ray structure of the CPs and what happens to them in aqueous solution during the first 25 ns. It was found that the N-terminal domain of CP was intrinsically unstructured. The principal component analysis (PCA) of the MD ensembles revealed components of collective atomic motions that promote the binding of viral RNA by the coat protein. It is interesting to note that there is a well-fixed short RNA fragment near the hinge region of CP chain A in the X-ray structure of TAV CP [4]. On the basis of this theoretical analysis, further suggestions are given for the better understanding of the mode of action of cucumoviral coat proteins.

## 2. Methods

The missing residues of the CMV and TAV CPs were modeled by the *ab initio* method using the ROSETTA [9–12] program suite. The completed, full-length coat protein structures were each subjected to 25 ns molecular dynamics simulations after the appropriate preparation. The GROMACS 3.3.1 [13–17] software package was used for running and evaluating the MD simulations.

### *2.1. The completion of the coat protein X-ray structures using ab initio model building*

#### *2.1.1. The preparation of the complete CP structures*

The Fny-CMV [3] coat protein C chain structure (PDB ID code: 1F15) and the TAV Blencowe strain [4] coat protein C chain (PDB ID code: 1LAJ) were used for the computer modeling of the full-length chain of the coat proteins because these polypeptide chains are the most complete of the three chains in the X-ray structures, lacking only the first 27 and 33 residues, respectively. Coat protein segments 1–47 were predicted *ab initio* using the ROSETTA program. This region was chosen because the N-terminal α-helix is located between residues 30 and 48 in the X-ray structures and the computed *ab initio* models could be easily joined to the existing X-ray structures. In both cases, a 100 decoy structures were generated with the ROSETTA program, and these were ranked into three and two clusters, respectively. The most adequate clusters were selected by visual evaluation, and the centroid structures of

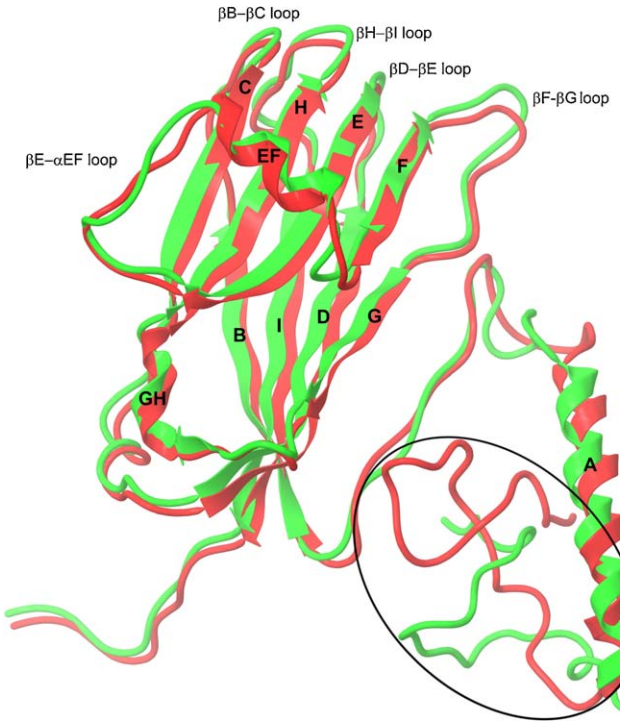


**Fig. 1.** The superimposed X-ray structures of CMV (green) and TAV (red) coat proteins completed by *ab initio* fragment modeling. The predicted regions are encircled. The sheets and helices are labeled according to Lucas et al. [4].

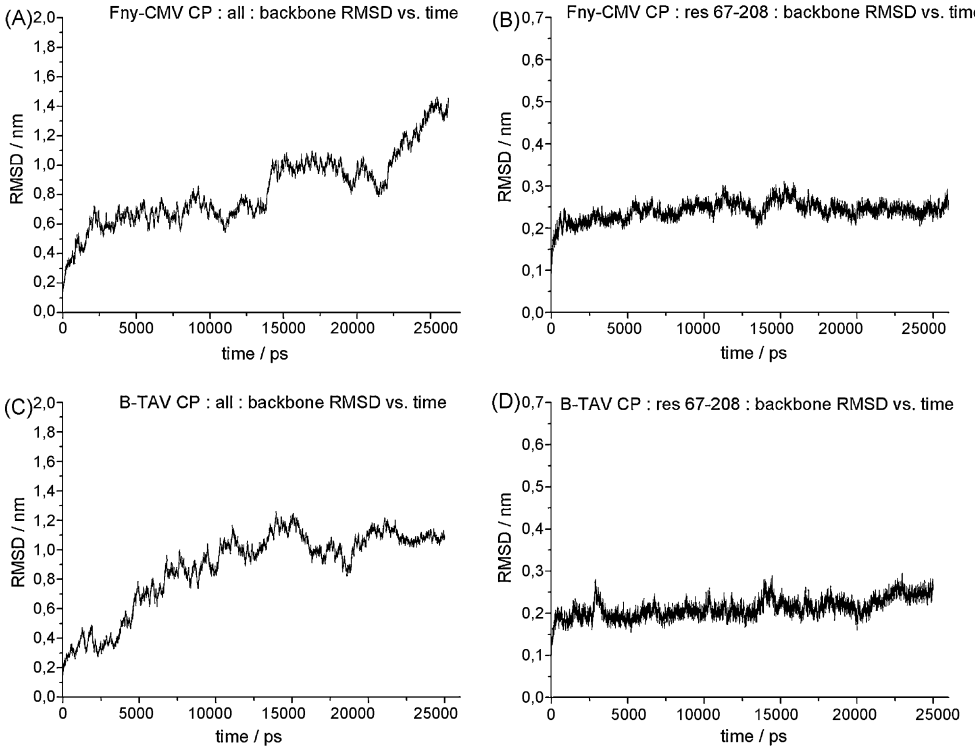


**Fig. 2.** The analyses of the CMV and TAV CP MD trajectories. The initial structures were taken as references when calculating the RMSD values. The trajectory plots were created in two ways. First, the RMSD values of the whole backbone the CMV CP and TAV CP chains were plotted as a function of time, after which the RMSD values of the core regions of the CMV CP and TAV CP from residues 67 to 208 were plotted.

the selected cluster were geometrically optimized. The first 29 and 38 residues of the predicted models were linked to the X-ray structures beginning with residues Ala30 (CMV) and Val39 (TAV), respectively (Fig. 1).

### 2.2. Molecular dynamics simulations

Before the MD simulations, both coat proteins were energy minimized using the GROMOS96 43b1 force field [18]. The MD

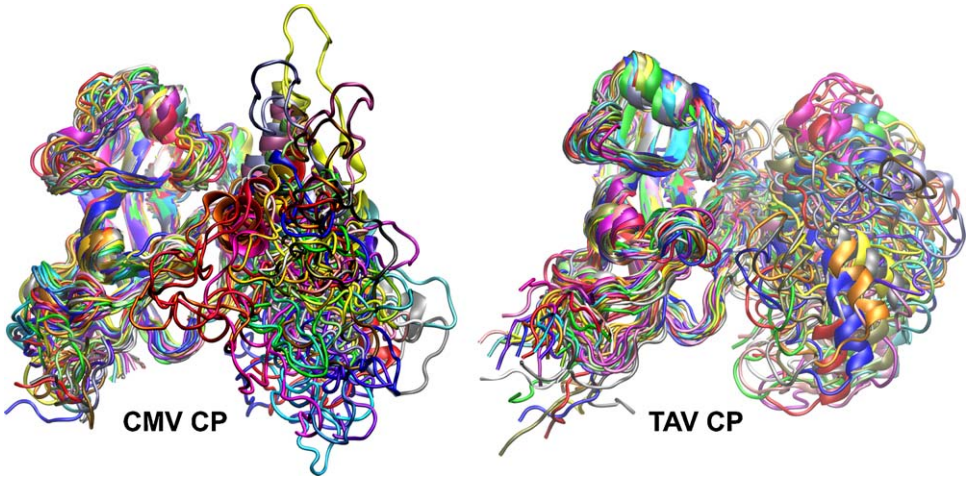


**Fig. 3.** The trajectories of the 25 ns molecular dynamics simulations. Snapshot structures taken every 500 ps were superimposed. It can be seen that the N-terminal domains and C-terminal tails of the coat proteins are intrinsically unstructured. The β-barrel core regions are rigid, while the loops between the β-sheets are quite flexible.

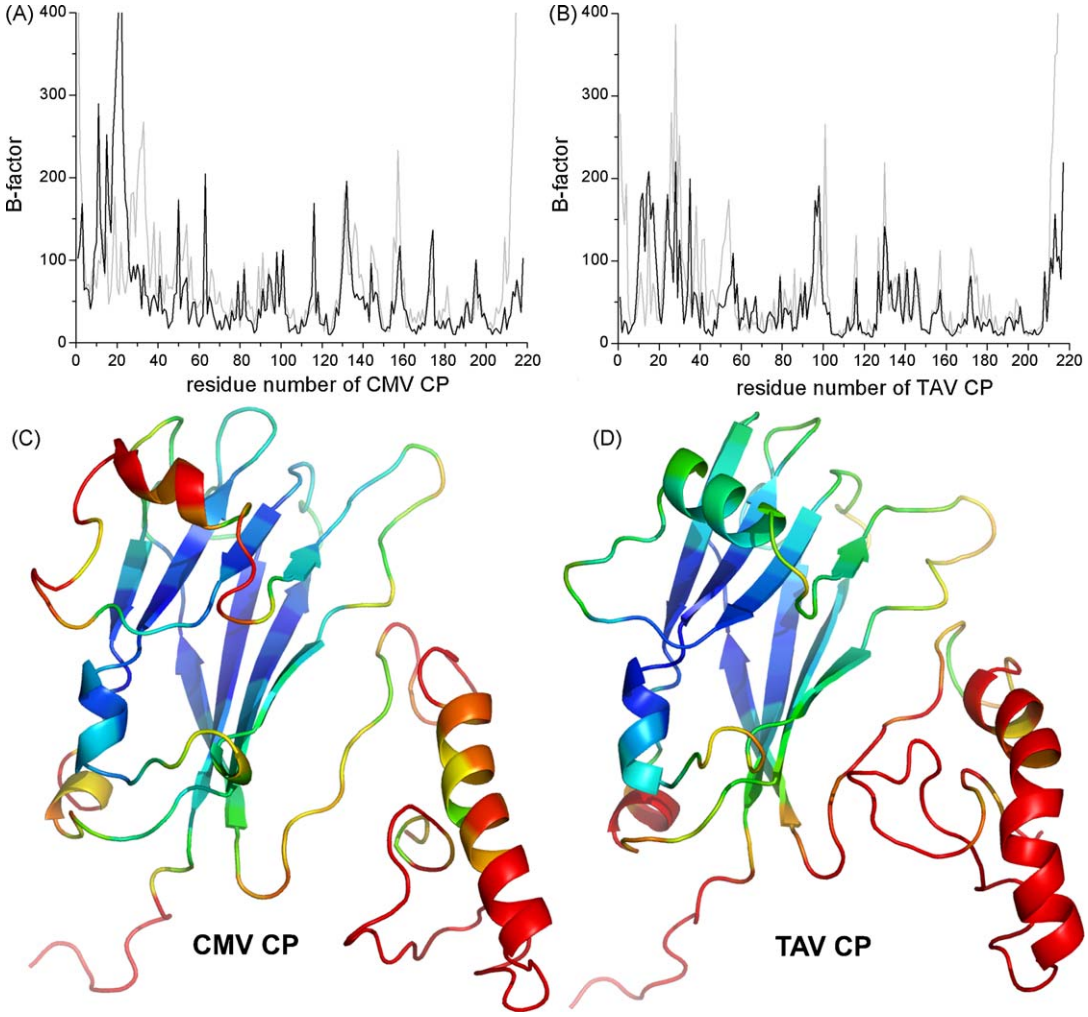


**Fig. 4.** The thermal factor analysis. The comparison of the averaged thermal factors (B-factor) per residue between the first 0.5 ns (0–0.5 ns) and the last 0.5 ns (24.5–25.0 ns) of the CMV (A) and TAV (B) CP molecular dynamics simulations. Light grey: the first 0.5 ns, black: the last 0.5 ns. The B-factors were plotted as a function of the coat protein residue number. The color visualization of the B-factors on the cartoon representation of CP models (C and D). The B-factor values are colored in rainbow succession. Blue is the most rigid and red is the most flexible.

simulations were performed on a HP xw9300 Linux workstation using four parallel 2.0 GHz AMD Opteron processor cores. The CPU time was around 50 h/ns.

### 2.2.1. The MD simulation of the CMV coat protein

The simulation system for the CMV coat protein was set up using a cubic box (5.271 nm × 5.343 nm × 5.610 nm), which includes the CP solvated with 18,889 SPC216 water molecules [19]. Twelve $Cl^-$ ions were added to the system to ensure electroneutrality, after which the system was filled up with 55 $Na^+$ and 55 $Cl^-$ ions to reach a salt concentration of 0.15 M. The simulation box contained a total of 18,767 water molecules. The initial system was energy minimized in two steps: first, only the position of the water molecules was minimized and then the potential energy of the whole system. The GROMOS96 53a6 force field [20] was used during the minimization. In the subsequent step on an NVT ensemble, a short 50 ps position-restrained MD simulation at 600 K was carried out using a Nose-Hoover thermostat to ensure the proper water settlement around the protein. After this, 0.5 ns of position-restrained MD simulation were performed on an NPT ensemble at 300 K and 1 bar using a Berendsen thermostat and pressure coupling. The whole system was then submitted to a strong energy minimization using the Polak-Ribiére conjugate gradient minimum searching method (PRCG). The minimal potential energy of the system was $-1.120082 \times 10^6$ kJ/mol, and the maximum force acting on 400 atoms was equal to 39.4647535 kJ $mol^{-1}$ $nm^{-1}$. Finally the unrestrained 25 ns MD simulation was initiated on the NVT ensemble. The major parameters of the simulation were as follows: The Newton equations of motion were integrated by the leap-frog algorithm using a 1 fs step size. Every 200th calculated structure (5 structures/ps sampling rate) and its physico–chemical parameters were saved into the output trajectory file. The electrostatics were handled by the Particle-mesh-Ewald (PME) method. A Berendsen thermostat ensured the temperature stability of the simulation at 300 K. The output data stream was saved in 500 ps segments for the easy handling of the output trajectory data.

### 2.2.2. The MD simulation of the TAV coat protein

The simulation for the TAV coat protein was carried out as described above for the CMV coat protein simulation. The simulation system was set up using a cubic box (5.443 nm × 5.446 nm × 5.499 nm) that included the CP molecule solvated by 19,733 SPC216 water molecules [19]. Ten $Cl^-$ ions were

**Table 1**
The comparison of the loop flexibility of coat proteins. Coat proteins have 218 aa in CMV and 217 aa in TAV because of a single deletion, Ser148, in CMV, so the numbering of loop residues differs by 1 after position 148.

| | Averaged thermal factors | | | |
|---|---|---|---|---|
| | CMV CP MD simulation | | TAV CP MD simulation | |
| | First 0.5 ns of the simulation | Last 0.5 ns of the simulation | First 0.5 ns of the simulation | Last 0.5 ns of the simulation |
| βB-βC loop (res 76–83) | 50.52 | 47.85 | 49.41 | 39.41 |
| βD-βE loop (res 113–119) | 38.93 | 64.08 | 47.51 | 29.12 |
| βE-αEF loop (res 129–136) | 113.44 | 106.08 | 88.30 | 76.59 |
| βF-βG loop<br>CMV CP (res 155–163)<br>TAV CP (res 154–162) | 90.75 | 48.90 | 43.19 | 30.08 |
| βH-HI loop<br>CMV CP (res 191–199)<br>TAV CP (res 190–198) | 44.05 | 49.83 | 33.18 | 26.71 |

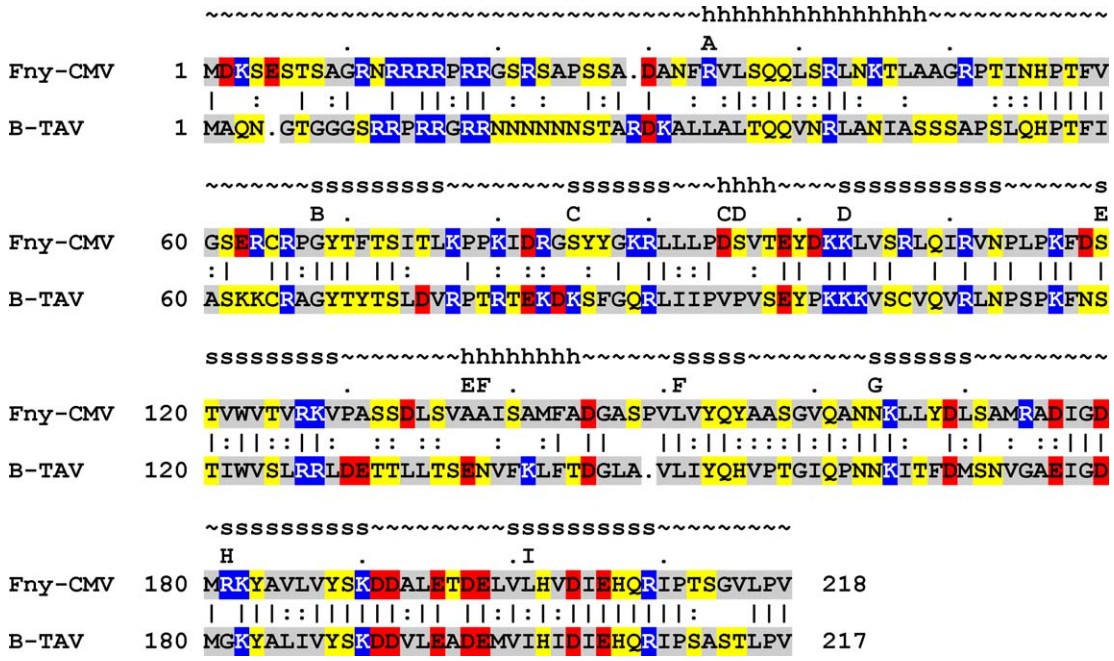


**Fig. 5.** The sequence alignment between Fny-CMV CP and B-TAV produced by the Gap program [39]. The secondary structure of the CPs is shown by the letters "s" for β-sheets and "h" for helices. The background of the sequence alignment reflects the chemical properties of the amino acids: acidic amino acids (aas) are red, basic aas are blue, polar aas are yellow and hydrophobic aas are grey.

**Table 2**
The results of the cluster analyses. The numbers represent the clusters found as a function of the given search conditions.

| Cut-off level (Å) | Full-length | | | | Core domain (residues 67–208) | | | |
|---|---|---|---|---|---|---|---|---|
| | CMV CP | | TAV CP | | CMV CP | | TAV CP | |
| | All atoms | Backbone | All atoms | Backbone | All atoms | Backbone | All atoms | Backbone |
| 10 | 2491 | 1120 | 2498 | 609 | 1417 | 164 | 1010 | 111 |
| 15 | 419 | 196 | 490 | 234 | 108 | 17 | 102 | 13 |
| 20 | 175 | 70 | 150 | 85 | 14 | 5 | 15 | 4 |
| 25 | 73 | 41 | 63 | 32 | 6 | 1 | 25 | 1 |

added to the system to ensure electroneutrality. After this, the system was filled up with 57 $Na^+$ and 57 $Cl^-$ ions to reach a salt concentration of 0.15 M. After the ion treatment, the simulation box contained 19,609 water molecules. The minimal potential energy of the system was $-1.1194795 \times 10^6$ kJ/mol and the maximum force acting on one atom was equal to 92.011078 kJ $mol^{-1}$ $nm^{-1}$ before the MD simulation.

### 2.3. The statistical evaluation of the MD simulations

The quality check on the MD simulations was performed using GROMACS applications. The root mean square fluctuations (RMSF) and deviations (RMSD) of atomic positions were plotted with "g_rmsf" and "g_rms", respectively. The salt bridges were calculated using "g_saltbr". The secondary structure distribution of the CPs was computed with "do_dssp" [21]. For the detailed investigation of the MD trajectories, they were subjected to cluster and covariance analyses. The "g_cluster" tool was used to cluster the CP structures over time. The gromos clustering method was chosen for all of the computations. Cluster analyses for both MD trajectories were computed for several combinations of clustering parameters. Four RMSD cut-off values (0.1 nm, 0.15 nm, 0.2 nm and 0.25 nm) were used to arrange neighboring structures. Clusters were also calculated for the backbone atoms and all protein atoms. The covariance matrices, eigenvectors and eigenvalues were constructed using the "g_covar" program. The essential dynamics analysis (PCA), projections of structures onto eigenvectors, analysis of eigenvectors for visualizing the main motions in the CPs and the calculation of the inner products between eigenvectors were carried out with the "g_anaeig" application. The molecular graphics were produced with PyMol (available at http://www.pymol.org), Maestro (Schrödinger, LLC) and VMD [22].

## 3. Results and discussion

### 3.1. Completed coat protein structures

The three-dimensional structures of the CMV and TAV CPs are almost completely identical due to the high amino acid sequence similarity (70.3%). It was thus worth using *ab initio* modeling to find the residues missing from the X-ray structures. The MD simulations provided additional information on the solution structures of the CMV and TAV coat proteins, and using the whole structure for modeling reflects the reality better. The superimposed structures of the coat proteins are illustrated in Fig. 1. The backbone root mean square deviation (RMSD) of the two structures was 4.78 Å and 2.68 Å for the core domain (residues 67–208).

### 3.2. MD simulations

#### 3.2.1. General characterization

The total energy was conserved for both simulations, and the temperature levels were stable when the trajectories of both MD simulations were monitored. The dynamic equilibrium state was not reached for the whole protein in 25 ns (Fig. 2). It can be seen that the structure of the N-terminal region from residues 1 to 66 and the C-terminal tail were highly disordered (Fig. 3). In addition, the N-terminal α-helix was partially unrolled. Analyzing the trajectories of the β-barrel core regions between residues 67 and 208, it can be concluded that the distortion of the intrinsically unstructured N-terminal region perturbs the dynamic equilibrium states (Fig. 2).

The CMV and TAV coat protein sequences were submitted to the IUPred server (Prediction of Intrinsically Unstructured Proteins; http://iupred.enzim.hu) [23,24]. The results of these queries were that the N-terminal regions of the CMV and TAV CPs from residues 1 to 43 were intrinsically unstructured. The theoretical predictions of the IUP regions confirmed the results of the MD simulations.

After visual investigation of the MD trajectories, averaged thermal factors (B-factor) per residue were calculated between the

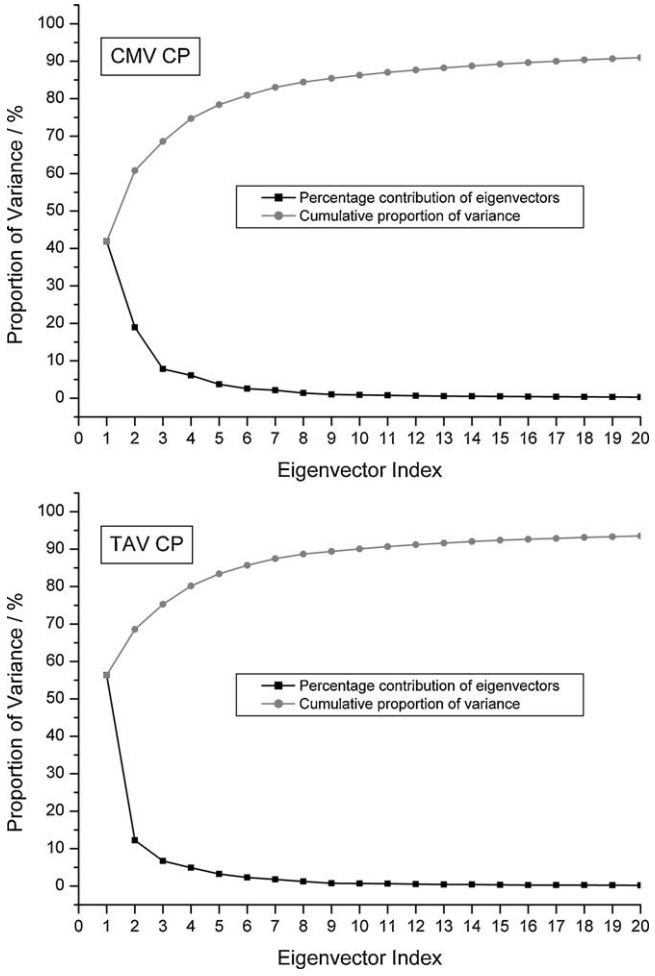


**Fig. 6.** The individual and cumulative proportion of the eigenvectors from the PCA on CMV and TAV CP MD ensembles. Only the contributions of the first 20 eigenvectors were plotted from the total variance.

first 0.5 ns (from 0.0 to 0.5 ns) and the last 0.5 ns (from 24.5 to 25.0 ns) intervals of the CMV and TAV CP MD simulations (Fig. 4A and B). A reduced index was created, which shows the overall difference in flexibility between the first and the last 0.5 ns MD simulations. This was calculated as follows: the averages of the averaged thermal factor per residue were computed for the first and last 0.5 ns time intervals and the average values belonging to the last 0.5 ns were subtracted from the average values of the first 0.5 ns simulations. In the case of the CMV CP, the value of this index was 21.21, while that of the TAV CP was 21.68. For these two simulations, it can be assumed that the overall differences in flexibility were almost identical and the overall values became smaller as the end of the simulations was approached, when the systems were in the quasi-equilibrium state previously discussed.

The flexibility indices for the loop regions were calculated by averaging the thermal factors of all atoms belonging to the given loop region (cf. Table 1 and Fig. 4). On the basis of the previously computed loop flexibility index values, it can be concluded that the βE-αEF loop was the most mobile of the five loops. The CMV CP loops were more flexible than the TAV CP loop regions. These loops are responsible for aphid vector transmission [25,26]. A larger number of symptom-determining amino acid residues play a role in these loops [8,27–32]. CMV has a much wider host plant range than TAV [2,33]. Virions of cucumoviruses make direct capsid interactions with the aphid vectors [34]. The greater flexibility of the CMV CP loop regions, which are located on the external surface of the virion, may help to initiate easier interaction formations within the stylets of aphids. However, it should be noted that the main characteristics of the loop regions are determined by their amino acid composition, providing specific electrostatic potential patterns [35,36].

The present MD simulations disproved an earlier hypothesis claiming that the flexibility of βE-αEF loops beginning with a proline residue is smaller than that of βE-αEF loops beginning with any other amino acid residue [37]. This study confirmed that the βE-αEF loop region of the CMV CP, starting with proline, was much more flexible than the same loop region in the TAV CP, where the first residue was asparagine (cf. Table 1).

So what causes the partially disordered property of the coat proteins? The physico–chemical reason for this disordered behavior is the rearrangement of several salt bridges in the N-terminal domain (data not shown), which contains several amino acids that (Fig. 5) are responsible for the interaction with viral RNA. In the absence of RNA, these salt bridges cannot form stable connections, and salt bridges are formed at random.

### 3.2.2. Cluster analysis

The partially disordered property of the CPs was completely verified by the cluster analysis of the MD trajectories, which revealed far fewer clusters for the core domain of the CP than when the whole protein sequence was included in the analysis. The results of the cluster analysis are presented in Table 2. Cluster centroids were also computed for each cluster. The number of clusters varied depending on the appropriate clustering cut-off value. In this particular case, the 20 Å cut-off value proved to be a reasonable choice (cf. Table 2). The β-sheets were remarkably stable during the simulation. The clusters of the core domain represented different structural classes of loop conformations. The solution structures of the CMV CP can be ranked into more clusters than those of the TAV CP because the CMV CP has more flexible loops.

### 3.2.3. Essential dynamics analysis

The general characterization of the MD systems only made it clear that the CPs examined did not behave in the same way during the simulations, as can be seen in Fig. 2. A principal component analysis was performed in order to obtain a clearer picture of the dynamical features of the coat proteins. In this way it became possible to characterize and compare the essential dynamics of both MD ensembles. In general terms, the first few principal components of a protein MD trajectory capture the essential modes of the main conformational motions [38]. In the present instance, the first three eigenvectors from the PCA on the CMV CP MD ensemble described 69% of the conformational motions, while this rate was 75% in the case of the TAV CP PCA. The subsequent principal components captured significantly smaller atomic

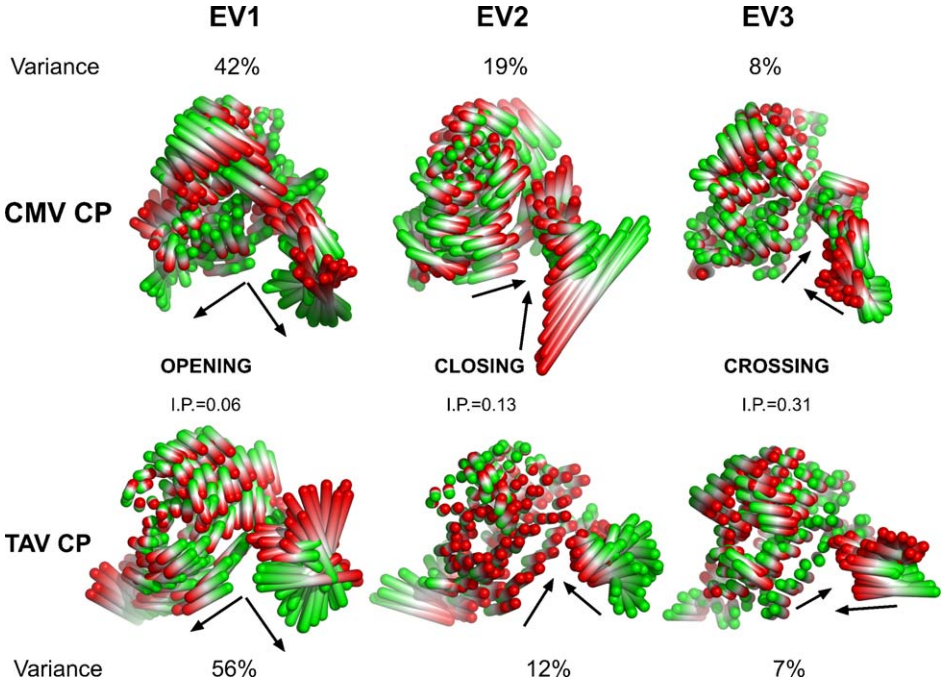


**Fig. 7.** The comparison of the first three principal components from the PCA of the CMV and TAV CP MD trajectories. Motions are illustrated as linear interpolations between the extreme projections of the structures onto the eigenvectors. The inner products (I.P.) among the first three eigenvectors from the CMV CP and TAV CP MD simulations are indicated. The shaded cylinders represent the movements of the carbon alpha atoms. Green is the starting state and red is the end state. The arrows indicate the conformational motion components of the RNA binding cleft.

fluctuations. It is clearly visible in Fig. 6 that the contribution of the remaining principal components to the total variance decreased after the first three eigenvectors. Therefore, accurate analyses of the principal components can be limited to the first three eigenvectors. Although the inner products (I.P.) between the corresponding PCA eigenvectors of the CMV and TAV CP ensembles had low values, similar trends can nevertheless be observed in the type of conformational motion (Fig. 7). Before discussing the main results, it is worth mentioning that the MD simulations on the coat proteins were run in the absence of RNA. The analysis focused on the RNA binding cleft of the CP, which is located between the hinge region and the N-terminal domain. The motion component described by the first eigenvector (EV1) performed an opening conformational movement in this RNA binding region as if it was "seeking for" the RNA molecule (Fig. 7). This principal component makes up the 42% and 56% of the total variance of the CMV and TAV CP MD ensembles, respectively. The second motion component (EV2) responsible for a smaller share of variance, showed a closing movement in this same region, while the third eigenvector (EV3), with the smallest variance contribution, described a crosswise collective motion. Another dynamic property of CPs was clearly visible from the comparison of the first three principal components: the loops of the CMV CP were more mobile than those of the TAV CP. The most flexible regions of the CMV CP were the βE-αEF loop and the EF-helix (cf. Figs. 4 and 7).

Conformational space sampling analyses were carried out for both simulations. The projection of the CMV and TAV CP trajectories onto the first two eigenvectors is shown in Fig. 8. It is immediately obvious that the CPs behaved entirely differently, which is justified by the partially disordered property of the N-terminal region. Some stability clusters could be noted during the CMV CP MD simulation, but the CP left the equilibrium state at the end of the simulation (cf. Fig. 2A and 8A). It is obvious from the projection of the TAV CP MD trajectory that the CP reached a relatively stable cluster during the last few nanoseconds of the MD simulation (cf. Fig. 2C and 8B). When the actual snapshot structures were assigned to the points of the 2D projections, it was observed that the N-terminal domains of both CPs performed mainly opening conformational motions during the simulations, but in opposite directions (inset of Fig. 8). The mutual positions of both N-terminal domains became almost orthogonal at the end of the simulations, which was well reflected by the low inner products of the eigenvectors.

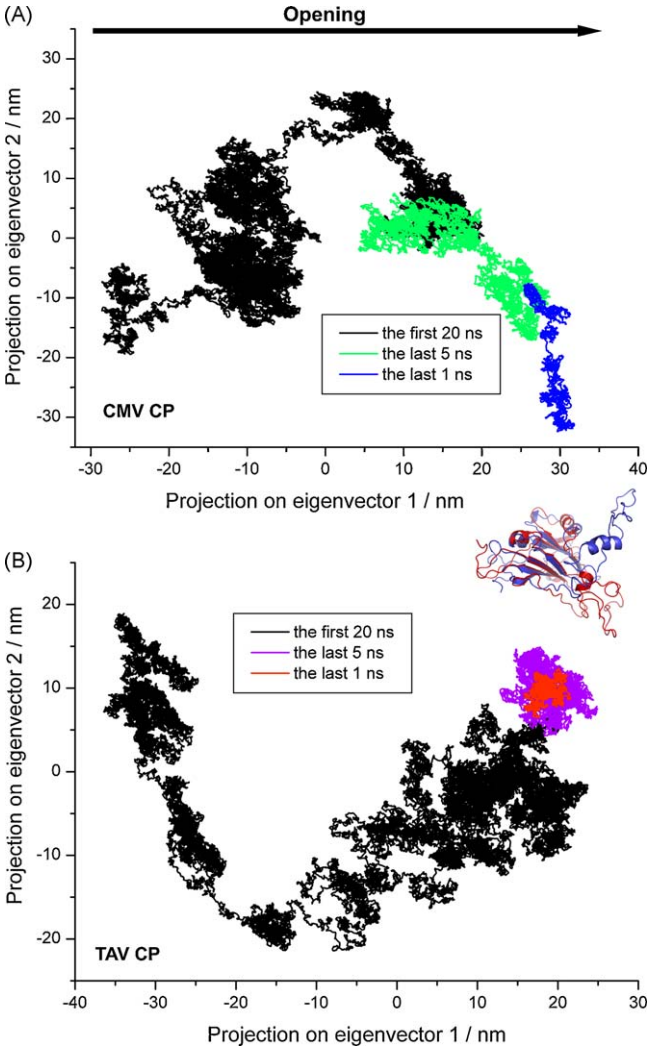


**Fig. 8.** The conformational space sampling. The projection of the CMV CP (A) and TAV CP (B) structures onto the plane encompassed by the first two eigenvectors from the PCA of the MD ensembles. The states of the simulations are colored according to the legends. The inset shows the CP structures at the end of the simulations in cartoon representation.

### 3.3. Possible role of the hinge region

There is a disulphide bridge between Cys64 and Cys106 in the TAV CP that is not present in the CMV CP. Disulphide bridges are located on the inner side of the virus particle and probably enhance the stability of the β-sheet. In a previous paper [8], it was suggested that there may be a new functional site on the surrounding region of amino acids 62 and 65. In addition, this cleft on the coat protein surface is responsible for an essential step in the cell-to-cell movement. The molecular graphics analysis (Fig. 9) of the three-dimensional structure of the TAV CP clearly indicates the localization of the disulphide bridge. A disulphide bond between Cys64 and Cys106 links β-sheet D and the hinge region. Comparing the average structures of the CMV and TAV CP, it can be concluded that the disulphide bridge in the TAV CP contributes significantly to the stabilization of the hinge region. However, the presence of this disulphide bond is not necessary for this RNA binding cleft to emerge. Basically, the hinge region of the CP is linked to the β-barrel core region by hydrogen bonds. There is a well-fixed RNA fragment (ApApA) near the hinge region of CP chain A in the X-ray structure [4] of TAV (Fig. 9). This hinge region may be the RNA recognition site of the CP.

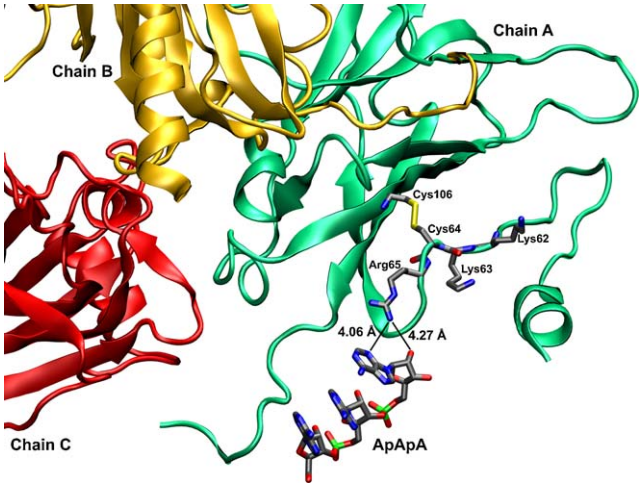


**Fig. 9.** The hinge region of the X-ray structure of the TAV CP. The CP subunit chains are in cartoon representation. Chains A, B and C are colored green, gold and red, respectively. The hinge region of Chain A is detailed, where the disulphide bridge and the ApApA RNA fragment are clearly visible.

## 4. Conclusions

This study provides insight into the dynamic behavior of the monomeric CMV and TAV coat proteins. The N-terminal domains of the cucumovirus CP structures have an intrinsically unstructured nature, making it easier for coat proteins to bind the RNA molecule, which is also partially disordered. Consequently, the position of the RNA fragment in the TAV X-ray structure [4], the previously detailed mutational experiments [8] and the type of conformational motion revealed by the PCA provide indirect evidence that the hinge region of CPs plays a crucial role in the recognition and binding of viral RNA. During viral RNA binding and virion assembly, the partially unfolded N-terminal helix must be refolded. The aggregation of these refolded amphiphilic helices may promote virion assembly and viral RNA binding to the inner side of the virion during assembly.

The MD simulations and PCA showed that the loop regions of the CMV CP are more flexible than those of the TAV CP. Owing to the greater flexibility of the CMV CP loop regions, they can adapt more easily to the mouth organ tissue of aphids.

The RNA binding domains of other RNA virus coat proteins may have similarly partially disordered RNA binding sites. The present results may help in elucidating the mode of action of cucumoviral coat proteins.

## Acknowledgments

Ákos Gellért is the recipient of a Bolyai János fellowship from the Hungarian Academy of Sciences. The licensing of the Schrödinger software package was funded from the AGRISAFE EU FP7 grant under agreement No. 203288. We thank Prof. Gábor Náray-Szabó for his helpful comments and critical reading of the manuscript. The authors are grateful to Mrs. Barbara Harasztos for revising the manuscript linguistically.

## References

[1] M.H.V. Van Regenmortel, C.M. Faquet, D.H.L. Bishop, E. Carstens, M. Estes, S. Lemon, J. Maniloff, M.A. Mayo, D. McGeoch, C.R. Pringle, R.B. Wicker, Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses, Academic Press, San Diego, 2000.
[2] P. Palukaitis, F. García-Arenal, Cucumoviruses, Advances in Virus Research 62 (2003) 241–323.
[3] T.J. Smith, E. Chase, T. Schmidt, K.L. Perry, The structure of Cucumber mosaic virus and comparison to cowpea chlorotic mottle virus, Journal of Virology 74 (16) (2000) 7578–7586.
[4] R.W. Lucas, S.B. Larson, M.A. Canady, A. McPherson, The structure of Tomato aspermy virus by X-ray crystallography, Journal of Structural Biology 139 (2) (2002) 90–102.
[5] J.C. Carrington, K.D. Kasschau, S.K. Mahajan, M.C. Schaad, Cell-to-cell and long-distance transport of viruses in plants, Plant Cell 8 (1996) 1669–1681.
[6] S.G. Lazarowitz, R.N. Beachy, Viral movement proteins as probes for investigating intracellular and intercellular trafficking in plants, Plant Cell 11 (1999) 535–548.
[7] I.B. Kaplan, L. Zhang, P. Palukaitis, Characterization of Cucumber mosaic virus. V. Cell-to-cell movement requires capsid protein but not virions, Virology 246 (2) (1998) 221–231.
[8] K. Salánki, Á. Gellért, E. Huppert, G. Náray-Szabó, E. Balázs, Compatibility of the movement protein and the coat protein of cucumoviruses is required for cell-to-cell movement, Journal of General Virology 85 (2004) 1039–1048.
[9] R. Bonneau, C.E. Strauss, C.A. Rohl, D. Chivian, P. Bradley, L. Malmstrom, T. Robertson, D. Baker, De novo prediction of three-dimensional structures for major protein families, Journal of Molecular Biology 322 (1) (2002) 65–78.
[10] R. Bonneau, J. Tsai, I. Ruczinski, D. Chivian, C. Rohl, C.E. Strauss, D. Baker, ROSETTA in CASP4: progress in *ab initio* protein structure prediction, Proteins: Structure, Function, and Genetics 5 (Suppl.) (2001) 119–126.
[11] K.T. Simons, I. Ruczinski, C. Kooperberg, B. Fox, C. Bystroff, D. Baker, Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins, Proteins: Structure, Function, and Genetics 34 (1) (1999) 82–95.
[12] K.T. Simons, C. Kooperberg, E. Huang, D. Baker, Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions, Journal of Molecular Biology 268 (1997) 209–225.
[13] E. Lindahl, B. Hess, D. van der Spoel, GROMACS 3.0: A package for molecular simulation and trajectory analysis, Journal of Molecular Modelling 7 (2001) 306–317.
[14] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation, Computer Physics Communications 91 (1995) 43–56.
[15] U. Essman, L. Perela, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle-mesh-Ewald method, Journal of Chemical Physics 103 (1995) 8577–8592.
[16] H.J.C. Berendsen, J.P.M. Postma, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, Journal of Chemical Physics 81 (1984) 3684–3690.
[17] S. Miyamoto, P.A. Kollman, SETTLE: an analytical version of the SHAKE and RATTLE algorithms for rigid water models, Journal of Computational Chemistry 13 (1992) 952–962.
[18] W.F. van Gunsteren, S.R. Billeter, A.A. Eising, P.H. Hünenberger, P. Krüger, A.E. Mark, W.R.P. Scott, I.G. Tironi, Biomolecular Simulation: The GROMOS96 Manual and User Guide, Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996.
[19] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, in: B. Pullman (Ed.), Intermolecular Forces, Reidel Publishing Company, Dordrecht, 1981, pp. 331–342.
[20] C. Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren, A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6, Journal of Computational Chemistry 25 (13) (2004) 1656–1676.
[21] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (12) (1983) 2577–2637.
[22] W. Humphrey, A. Dalke, K. Schulten, VMD-visual molecular dynamics, Journal of Molecular Graphics 14 (1996) 33–38.
[23] Z. Dosztányi, V. Csizmók, P. Tompa, I. Simon, The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins, Journal of Molecular Biology 347 (4) (2005) 827–839.
[24] Z. Dosztányi, V. Csizmók, P. Tompa, I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, Bioinformatics 21 (2005) 3433–3434.
[25] B. Chen, R.I.B. Francki, Cucumovirus transmission by the aphid Myzus persicae is determined solely by the viral coat protein, Journal of General Virology 71 (1990) 939–944.
[26] K.L. Perry, L. Zhang, P. Palukaitis, Amino acid changes in the coat protein of Cucumber mosaic virus differentially affect transmission by the aphids Myzus persicae and Aphis gossypii, Virology 242 (1) (1998) 204–210.
[27] D. Szilassy, K. Salánki, E. Balázs, Stunting induced by Cucumber mosaic cucumovirus-infected Nicotiana glutinosa is determined by a single amino acid residue in the coat protein, Molecular Plant-Microbe Interaction 12 (12) (1999) 1105–1113.
[28] M. Suzuki, S. Kuwata, J. Kataoka, C. Masuta, N. Nitta, Y. Takanami, Functional analysis of deletion mutants of Cucumber mosaic virus RNA3 using an in vitro transcription system, Virology 183 (1991) 106–113.
[29] S.M. Wong, S.S.C. Thio, M.H. Shintaku, P. Palukaitis, The rate of cell-to-cell movement in squash of Cucumber mosaic virus is affected by sequences of the capsid protein, Molecular Plant-Microbe Interactions 12 (1999) 629–632.
[30] K.H. Ryu, C.H. Kim, P. Palukaitis, The coat protein of Cucumber mosaic virus is a host range determinant for infection of maize, Molecular Plant-Microbe Interactions 11 (5) (1998) 351–357.
[31] S. Liu, X. He, G. Park, C. Josefsson, K.L. Perry, A conserved capsid protein surface domain of Cucumber mosaic virus is essential for efficient aphid vector transmission, Journal of Virology 76 (2002) 9756–9762.
[32] K.L. Perry, L. Zhang, M.H. Shintaku, P. Palukaitis, Mapping determinants in Cucumber mosaic virus for transmission by *Aphis gossypii*, Virology 205 (2) (1994) 591–595.
[33] P. Palukaitis, M.J. Roossinck, R.G. Dietzgen, R.I. Francki, Cucumber mosaic virus, Advances in Virus Research 41 (1992) 281–348.
[34] R. Hull, in: R. Hull (Ed.), Matthew's Plant Virology, Elsevier Academic Press, 2004, p. 497.
[35] L.F. Pacios, F. García-Arenal, Comparison of properties of particles of Cucumber mosaic virus and Tomato aspermy virus based on the analysis of molecular surfaces of capsids, Journal of General Virology 87 (2006) 2073–2083.
[36] V.D. Bowman, E.S. Chase, A.W.E. Franz, P.R. Chipman, X. Zhang, K.L. Perry, T.S. Baker, T.J. Smith, An antibody to the putative aphid recognition site on Cucumber mosaic virus recognizes pentons but not hexons, Journal of Virology 76 (2002) 12250–12258.
[37] Á. Gellért, K. Salánki, G. Náray-Szabó, E. Balázs, Homology modelling and protein structure based functional analysis of five cucumovirus coat proteins, Journal of Molecular Graphics and Modelling 24 (2006) 319–327.
[38] A. Amadei, A.B.M. Linssen, H.J.C. Berendsen, Essential dynamics of proteins, Proteins Structure Function and Genetics 17 (4) (1993) 412–425.
[39] S.B. Needleman, C.D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins, Journal of Molecular Biology 48 (3) (1970) 443–453.