

# The sequence and model structure analysis of three Polish peanut stunt virus strains

Aleksandra Obrepalska-Stepłowska ·  
Katarzyna Nowaczyk · Marta Budziszewska ·  
Anna Czerwoniec · Henryk Pospieszny

Received: 10 October 2007 / Accepted: 19 November 2007 / Published online: 30 November 2007  
© Springer Science+Business Media, LLC 2007

**Abstract** Peanut stunt virus (PSV) belongs to the *Cucumovirus* genus of the family *Bromoviridae* and is widely distributed worldwide, also in Poland. PSV is a common pathogen of a wide range of economically important plants. Its coat protein (CP), similarly as in other viruses, plays an important role in many processes during viral life cycle and has great impact on the infectivity. In this study, we present the results of sequence–structure analysis of CP derived from three Polish strains of PSV: PSV Ag, G, and P. Sequences were determined using RT-PCR amplification followed by sequencing and

compared with each other and also with CP from other known PSV viruses. We analyzed their phylogenetic relationship, based on CP sequence, using bioinformatic tools as well as their spatial model using homology-modeling approach with combination of ROSETTA algorithm for de novo modeling. We compared our model with those recently obtained for other cucumoviruses including PSV-Er. Our results have shown that all Polish strains probably belong to the first subgroup of PSV viruses. Homology level between strains Ag and G proved very high. Using theoretical modeling approach we obtained a model very similar to the one resolved previously with the differences caused by slightly different amino acid sequence. We have also undertaken an attempt to analyze its distant regions; however, results are not unequivocal. Analysis of symptoms and their correlation with specific amino acid position was also performed on the basis of results published elsewhere. The definite interpretation is impeded by the presence of satellite RNAs in Ag and P strains modulating symptoms' severity, though.

The nucleotide sequence data reported in this paper have been submitted to the GenBank nucleotide sequence database and have been assigned the accession numbers: EF693944 for PSV-Ag, EF535261 for PSV-G, and EF535260 for PSV-P strain.

A. Obrepalska-Stepłowska (✉) · K. Nowaczyk ·  
M. Budziszewska  
Interdepartmental Laboratory of Molecular Biology, Plant  
Protection Institute, W. Węgorza 20 Street, 60-318 Poznan,  
Poland  
e-mails: ao.stepłowska@ior.poznan.pl; olaob@o2.pl

K. Nowaczyk  
e-mail: k.nowaczyk@ior.poznan.pl

M. Budziszewska  
e-mail: M.Budziszewska@ior.poznan.pl

A. Czerwoniec  
Bioinformatics Laboratory, Institute of Molecular Biology and  
Biotechnology, Adam Mickiewicz University, Umultowska 89  
Street, 61-614 Poznan, Poland  
e-mail: anna.czerwoniec@amu.edu.pl

H. Pospieszny  
Department of Virology and Bacteriology, Plant Protection  
Institute, W. Węgorza 20 Street, 60-318 Poznan, Poland  
e-mail: h.pospieszny@ior.poznan.pl

**Keywords** Viral coat protein · Phylogeny ·  
Homology modeling

## Abbreviations

aa	amino acid(s)
bp	base pair(s)
nt	nucleotide
e-value	expectation value
CP	coat protein
ARM	arginine-rich motif
MSA	multiple sequence alignment
PBD	protein data bank
SatRNA	satellite RNA

## Introduction

Peanut stunt virus (PSV) belongs to the genus *Cucumovirus* whose members are transmitted by aphids and have the widest host range (over 1,000 species) among all known viruses in the world. In addition to PSV, *Cucumovirus* genus comprises also *Cucumber mosaic virus* (CMV) and *Tomato aspermy virus* (TAV). Apart from peanut, PSV also infects many other economically important crops such as pea (*Pisum sativum* L.), bean (*Phaseolus vulgaris* L.), yellow lupine (*Lupinus luteus* L.), etc [1, 2].

PSV has been reported in many countries worldwide. Strains of PSV have been classified into two subgroups: I (eastern) and II (western), based on the homology of the RNA3 nucleotide (nt) sequences [3]. Then, the third group has been distinguished, based on lower homology level to the previous two groups [4]. The genome of PSV is tripartite and positive sense ssRNA, where RNA1 and RNA2 encode replicase complex. RNA3 carries the information of movement protein (3a) and its negative strand is transcribed into fourth subgenomic RNA (RNA4), which is the messenger for coat protein (CP). PSV virions may also contain a fifth component, designated satellite RNA (sat-RNA) that could modulate severity of disease symptoms (reviewed in [5, 6]).

The molecular weight of CP of cucumoviruses is 24 kDa and 180 of its copies contribute to viral capsid. Primary function of the CP is encapsidation of viral nucleic acids; however, the role of CP is more complex and its importance during different phases of viral infection had been confirmed. CP of some cucumoviruses (CMV) is known to be involved in symptom determination and viral spread in plants [7]. It also contains determinants important for transmission in aphids [8]. The CP of one of the *Cucumovirus* members (CMV) has also been reported to participate in cell-to-cell and long distance movement together with movement protein [9]. The protein sequence of CP seems to be important for the infectivity of the virus, as substitution of some amino acids (aa) can change the infectious potential of the virus, including conversion from non-infectious into infectious [10, 11]. Some correlation between specific aa residues and symptoms character, in the various host species, were also reported. For example, it has been found that asparagine at position 193 in CMV-R is responsible for strong stunting response whereas lysine at the same position in CMV-Trk7 induces green mosaic [10].

The N-proximal region of CP in *Bromoviridae* family has been found to be arginine rich (ARM—arginine-rich motif) and to possess the capacity to bind genomic RNA. This interaction is important for encapsidation process and pathogenicity of the virus [12].

In this report we have undertaken attempts to analyze nt and aa sequences of CP in Polish strains of PSV, their

phylogenetic relationships with other known PSV strains, and structural model of CP.

## Materials and methods

### PSV strains and inoculation trials

In this study three isolates of PSV were analyzed: from celery PSV-Ag, pea—PSV-G, and yellow lupin PSV-P. They were maintained and propagated in *Pisum sativum* and *Phaseolus vulgaris*.

To prepare inoculums for mechanical transmission, the infected pea plants, showing systemic symptoms, were homogenized in 0.05 M phosphate buffer, pH 7.5 and the sap was used for inoculation of carborund-dusted leaves of the plants. The test plants (Table 1) were grown in a greenhouse at 18–24°C, with a 14 h light period. Plants were observed for symptoms development for 4 weeks after inoculation. The experiment was conducted in triplicate and for each trial at least three plants of each species were used.

### Purification of viral particles, RNA extraction, and RT-PCR amplification

Purification of the virus and RNA extraction were performed as described previously [13, 14]. Integrity and length of the extracted RNA was checked by separation on agarose denaturing gel stained with ethidium bromide and visualized under UV light.

Reverse transcription of the viral RNA was carried out using Superscript III Reverse Transcriptase (Invitrogen, Warsaw, Poland) with primer complementary to 3' UTR region of RNA3, according to manufacturer's instruction. Amplification was performed with cDNA obtained in previous step as a template. PCR reaction was carried out with a total volume of 20 µl with specific primers complementary to intergenic sequence between 3a and CP protein genes (upper primer) and to 3' UTR region of RNA3 (lower primer), as described previously [15].

Reaction was carried out in a DNA Mastercycler Personal (Eppendorf, Poznan, Poland). Products of amplification were separated in agarose gel with addition of ethidium bromide and visualized under UV light.

### DNA cloning and sequencing

DNA fragments obtained in the PCR reaction were isolated from the agarose gel with QiaExII Gel Extraction Kit (QIAGEN, Wrocław, Poland) and then cloned into pGEM-T-Easy Cloning Vector System (Promega, Straszyn, Poland), according to the manufacturer's instructions.

**Table 1** Symptoms of infection after inoculation with three Polish isolates of PSV for different plant hosts

Plant species	PSV-P		PSV-G		PSV-Ag	
	Local infection	Systemic infection	Local infection	Systemic infection	Local infection	Systemic infection
<i>Chenopodium amaranticolor</i>	Lchl (Lnl)	–	Lchl	Mo, Ma, St	Lchl (Lnl)	–
<i>Spinacia oleraceae</i>	Lchl	l	Lchl	L	Lchl	Mo, Ma
<i>Gamphrena globosa</i>	Lrdl	Mo	Lrdl	–	Lrdl	Mo, Ma
<i>Arachis hypogea</i>	–	–	–	–	–	Schl, Snl, Ma
<i>Phaseolus vulgaris</i>	Lchl	Mo	Lchl	Sn, Mo, Ma, St	Lchl	Mo (Sn)
<i>Pisum sativum</i>	–	Mo	Lchl	Mo, Ma, St	–	Mo
<i>Datura stramonium</i>	Lchl	l	Lchl	Mo	Lchl	Mo, Ma
<i>Nicotiana paniculata</i>	Lchl	–	Lnl0	–	Lnl	–
<i>Apium graveolens</i>	Lchl	–	n.t.	n.t.	Lchl	Schl, Vc, Chl
<i>Nicotiana rustica</i>	Lchl, Lnl	l	Lchl	Lchl	Lchl	Schl

Lchl, local chlorotic lesions; Lnl, local necrotic lesions; Lrdl, local reddish lesions; Mo, mosaic; Ma, malformation; Chl, chlorosis; Sn, systemic necrosis; St, stunt; Vc, vein clearing; Schl, systemic chlorotic lesions; Snl, systemic necrotic lesions; l, latent infection (presence of virus was confirmed by ELISA tests); –, no symptom, no virus, n.t., not tested; (), symptoms (infection) appeared occasionally

Plant species were chosen on the basis of differentiating symptoms on them caused by strains PSV-Ag and PSV-P

*Escherichia coli* strain TG1 competent cells were transformed by electroporation in Micro Pulser electroporation system (Bio-Rad, Warsaw, Poland). Recombinant plasmids were isolated using QIAprep Spin Miniprep Kit (QIAGEN, Wrocław, Poland) and then automatically sequenced (IBB Sequencing Service, Warsaw, Poland).

#### Phylogenetic analyses

The nt and aa sequences of the PSV gene encoding CP were analyzed. The CP sequences of Polish isolates, Ag, G, and P, were compared with those of other PSV viruses and one strain of TAV virus from GeneBank. The sequences were compared using GeneDoc software [16]. Multiple sequence alignments (MSA) of sequences encoding CP were performed using CLUSTAL X [17]. After an initial comparison and predetermining of the variable and conservative regions, phylogenetic analyses were carried by using MEGA 3.1 [18] with ME method (Minimum Evolution, 1,000 repetitions) and with bootstrap test. Phylogenetic trees were drawn from the CP sequence of TAV virus, strain KC-TAV, as an outside group, and visualized using MEGA 3.1 software.

The CP nt and aa sequences were also compared with other known PSV sequences using BLASTN and BLASTP algorithms, respectively (Blast server, NCBI). Additionally, Polish isolates were compared against each other using BLAST2SEQ [19].

#### Theoretical modeling of Polish PSV CPs

To model the PSV-P, PSV-Ag, and PSV-G CP structure, we used the combination of the “Frankenstein’s Monster”

approach for template-based modeling [20] with the ROSETTA algorithm for de novo modeling [21]. Structural homologs were recognized and target-template alignments were generated using a fold-recognition method (GeneSilico MetaServer) [22]. The target alignments were converted into preliminary models using MODELLER [23] and those models were evaluated according to knowledge-based potentials (COLORADO3D) [24]. After superimposing the best models, we constructed hybrid models and used them to guide modifications of the original target-template alignments. The iterative model building, evaluation, and realignments were used as a refinement of models. For regions (1–27 and 203–216) with no corresponding structure among the templates identified by fold-recognition methods, we attempted de novo modeling using the ROSETTA algorithm. Typically, hundreds of thousand of decoys were generated and clustered to identify the most representative low-energy conformations. Models were selected according to the average energy clusters, size, density, and visual evaluation of the full-atom structures. The final hybrid models were “refined” by using MODELLER to optimize the bond lengths and angles. Final models were evaluated using PROQ server [25, 26] and V3D [24]. Mapping of the electrostatic potential on protein surfaces was calculated with APBS (Adaptive Poisson-Boltzmann Solver) [27].

## Results

#### Symptoms description

The symptoms observed on the tested plants are summarized in the Table 1. In general, we can notice that PSV-G

caused the most severe symptoms in *P. vulgaris* and *P. sativum* like stunting, malformation, and local chlorotic lesions, while PSV-P and -Ag caused mosaic and local chlorosis in these hosts. Our greenhouse observations also showed that PSV-Ag is more virulent, causing necrotic lesions in some plant hosts, than PSV-P isolate in spite of the presence of satellite in both strains. Additionally, symptoms varied depending on the host species.

#### CP sequences analysis and phylogenetic study

Sequencing of genes encoding Cps has shown that they consist of 654 nt in case of PSV-Ag and PSV-G strains and of 651 nt in case of PSV-P and code 217 and 216 aa proteins, respectively. Their sequences are available in Genbank with accession numbers: EF693944 for Ag, EF535261 for G, and EF535260 for P strain.

The comparison of obtained sequences with five other CP genes of PSV, available in GenBank, tentatively classifies Polish sequences to the first subgroup represented by strains J and Er (Table 2). While Ag and G strains share the highest homology level with Japanese strain J (87% of nt and 85% of aa identities); strain P has the greatest affinity to Er (86% and 85%, respectively).

The differences with subgroup III and its representative strain Mi are significant, while subgroup II and W strain are the most distant in regard of studied Polish strains.

Sequence analysis of the Cps of PSV isolates showed very high homology level between these sequences. N-proximal terminus of CP is very rich in arginine residues, which is characteristic for other *Bromoviridae* containing ARM [28, 29] (Fig. 1).

According to Hajimorad et al. [31], PSV strains can be considered as the separate subgroups within the same

species when their nt sequence identity is in the range of 70–80%. To the same subgroup belong the strains with the nt sequence identities greater than 90%. Coat proteins of analyzed nt sequences display less than 90% identity with the most similar strains of I subgroup (Table 2), but they can not be considered as a separate subgroup, because that homology is much higher than 70–80%.

The similarity between CP sequences of strains Ag and G is striking. They share 99% identity (Table 3). The characteristics of other corresponding genes are similar (unpublished data).

Phylogenetic analyses confirmed that Polish strains of PSV belong to one subgroup and they had common ancestor in the past (Fig. 2). They create together with other members of I subgroup a distinct cluster. Close relationship of CP from PSV strains, Ag and G, with Japanese strain J as well as PSV-P with PSV-Er was reinforced by the obtained phylogenetic tree. The tree is supported by high bootstrap values.

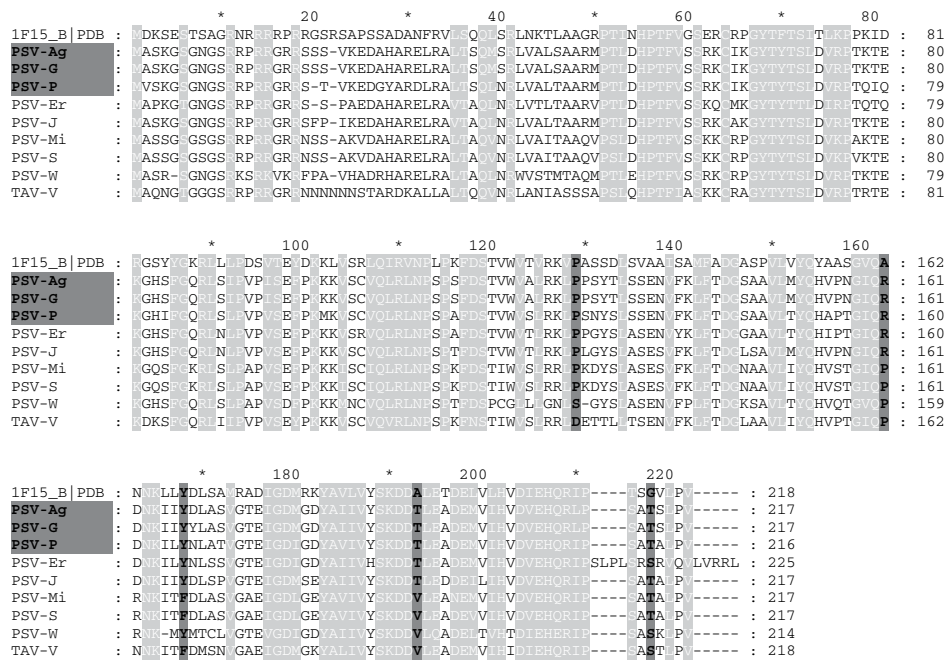
#### Analysis of the model

Amino acid residues responsible for infectivity of the virus have been identified in the core of the CP of three Polish isolates of PSV. Their conformation has been determined and surface potential was assessed. We have tried to compare data obtained after analysis of the model and greenhouse experiments with results of previous research, connecting the conformation of certain amino acids with infectivity of the virus [10, 11, 30].

For each of the three analyzed PSV isolates, the core of CP was modeled using homology-modeling approach. Modeling was carried out on the basis of the most closely related structures of *Cucumber Mosaic Virus* (1F15 in PDB), chain B (e.g., SAM score:  $1.2e-54$ , MGENTHREADER score: 0.944, 3DPSSM score:  $8.7e-08$ ) and *Tomato Aspermy Virus* (1LAJ in PDB), chain A (e.g., SAM score:  $4.8e-40$ , MGENTHREADER score: 0.867, 3DPSSM score: 0.024). Because N- and C-termini (first 27 and last 13 aa) do not have their counterparts in identified templates, they have been modeled de novo. The core of the protein consists of eight  $\beta$ -sheets and five  $\alpha$ -helices and, therefore, is in accordance with the characteristic motif of other *Bromoviridae* coat proteins [32]. The de novo-modeled N-terminal region comprises one  $\alpha$ -helix (aa 20–29); the conformation of aa 1–20, the sequence very rich in arginine residues, remains impossible to determine unequivocally. At the C-terminal region of the protein, de novo modeling resulted in obtaining the model with elongated last  $\beta$ -sheet (aa 204–207). Five of the last aa (210–214) could adopt a conformation of a short helix, although these results were not unambiguous. Results of

**Table 2** Sequence homology of CPs from Polish strains of PSV virus and other PSV strains representing specific subgroups

Basis of comparison		Strain J (subgroup I) (%)	Strain W (subgroup II) (%)	Strain Mi (subgroup III) (%)
<i>Polish strains</i>				
Ag	nt identities	87	76	77
	aa identities	85	66	76
	aa similarities	95	81	91
G	nt identities	87	76	77
	aa identities	85	66	75
	aa similarities	95	81	91
P	nt identities	85	75	76
	aa identities	82	63	75
	aa similarities	92	76	89



**Fig. 1** MSA of the CP of Polish PSV isolates and other members from *Cucumovirus* genus, available in databank, generated with CLUSTALX and visualized with GeneDoc. Amino acid conservation of 80% and higher is marked by light gray shading. Amino acid residues colored dark gray and indicated in bold are probably responsible for the symptoms evolving (according to [30]). 1F15

sequence represents the aa sequence of the CMV-Fny isolate with resolved X-ray structure. Then Polish strains as well as other known PSV CP aa sequences from Polish strains as well as TAV-V isolate as an outgroup are shown. Sequence names of Polish strains are colored dark gray and indicated in bold

**Table 3** Sequence homology between Polish PSV strains

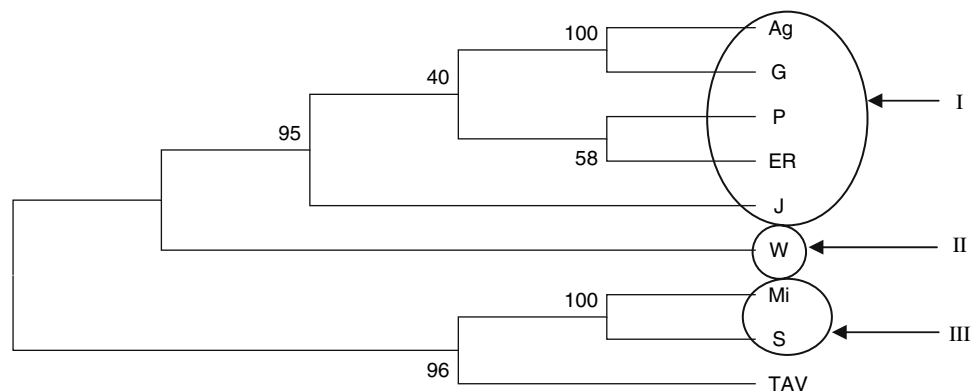
nt identities	aa identities and similarities:		
	Ag	G	P
Ag	–	99%/99%	85%/95%
G	99%	–	85%/95%
P	87%	87%	–

Nucleotide identities and aa identities and similarities obtained in BLAST2SEQ comparisons are presented

evaluation of all final models were good: for helical conformation of N-terminal region as well as for disordered N-terminus (Table 4 and Fig. 3).

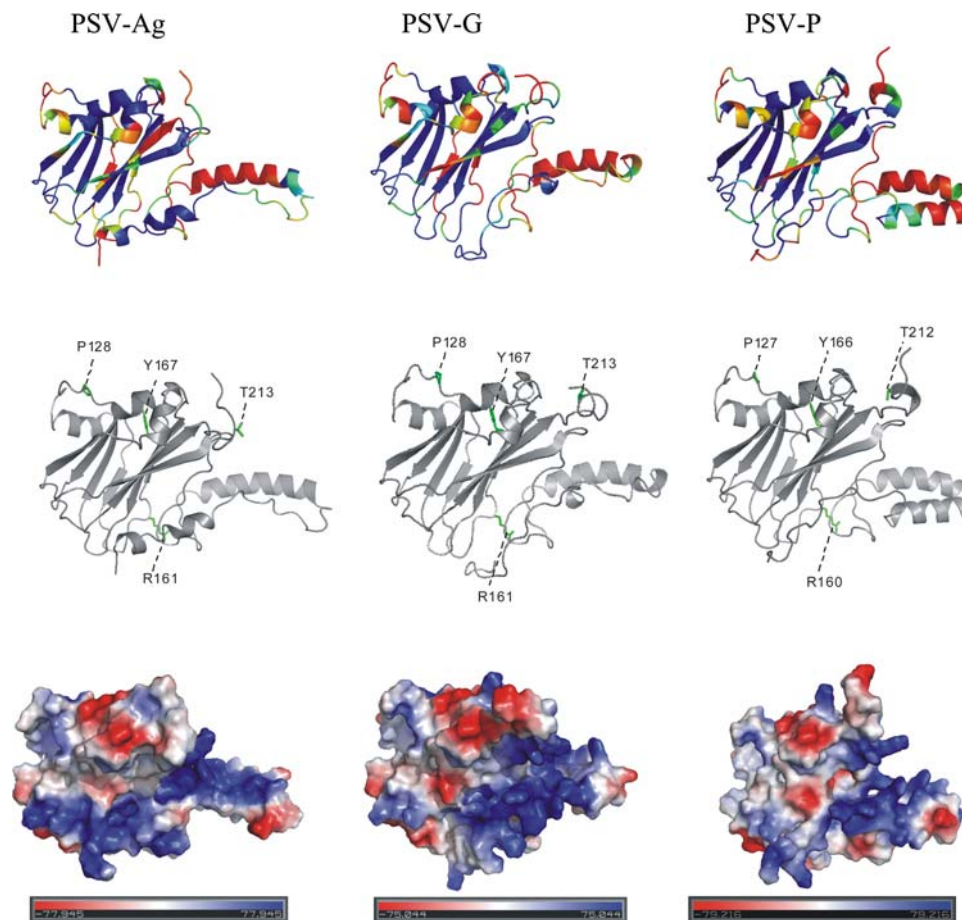
On the basis of the theoretical model and experimental data, aa playing crucial role in infection process [10, 30] were identified and the conformation of their surroundings compared to literature data. In the reference publications the following aa were specified: 129 proline/leucine, 162 threonine/alanine, 168 cysteine/tyrosine, 193 asparagine/lysine (as well as mutants of lysine residue: K193S and K193N), and 214 glycine/arginine in CMV. In the case of Polish strains of PSV in aforementioned localization we indicated Pro at 127 position in -P and 128 in -Ag and -G; Arg at 160 in -P and 161 in -Ag and -G; Tyr at 166 in -P and 167 in -Ag and -G; Thr at 191 in -P and 192 in -Ag and -G; and Thr at 213 in -P, and 212 in -Ag and -G. The

**Fig. 2** Phylogenetic tree of PSV strains from Poland, Japan, USA, and China with TAV strain as an outgroup, obtained on the basis of CP sequences. Subgroups I–III are indicated by circles



**Table 4** Model quality assessment using PROQ sever

PSV strain	PROQ LG score	PROQ MaxSub score	PROQ ranges of quality		
			Correct	Good	Very good
Ag	2.459	0.277	LGscore > 1.5	LGscore > 3	LGscore > 5
G	3.062	0.126	MaxSub > 0.1	MaxSub > 0.5	MaxSub > 0.8
P	2.749	0.125			



**Fig. 3** Models of PSV-Ag (left column), PSV-G (middle column), and PSV-P (right column). Coordinates in the PDB format, PyMol session files (allowing for displaying colors, surfaces etc. in three dimensions). All representations of a given protein are shown in the same orientation and scale. The first upper row presents models in the ribbon representation, colored according to the predicted local deviation from the real structure (i.e., the predicted error of the model), as calculated by Verify3D. Blue indicates low predicted deviation of C $\alpha$  atoms down to 0 Å, red indicates unreliable regions

with deviation >5 Å, green to orange indicate intermediate values. The second row shows models in the ribbon representation, with ligands and selected functionally important aa residues shown in the wireframe representation and labeled (P127/128; R160/161; Y166/167; T212/213). The third, bottom row shows proteins in the surface representation, colored according to the distribution of the electrostatic surface potential calculated with APBS (see Methods). Blue indicates positively charged regions and red indicates negatively charged regions

different positions result from deletion of one aa in CP sequence in PSV-Ag and -G, and two in case of PSV-P. The conformation of potentially important aa and the surface potential is shown in Fig. 3 (middle panel). Identified important aa are located near the molecular surface and moderately accessible, apart from T191/192, which is buried inside the protein.

## Discussion

Coat proteins of viruses play many crucial functions in their life cycle and infection process. They are also important for phylogeny studies. In our work we have tried to analyze three Polish PSV strains' CP using bioinformatic tools.

Our research has shown that the identity of studied strains accounts to almost 90% of the strains representing the first subgroup of PSV. Although PSV viruses may be qualified to the same subgroup when they share more than 90% homology [31], there are no reasons to qualify our strains to completely different subgroups, though. Identity between Polish strains is by far greater than 80%, which exceeds requirements for classifying viruses to separate subgroups.

Similarity of nt and aa sequences of CP from Ag and G strains is very high. Noticeable is the fact, that although their CP sequences are almost identical, symptoms in tested plants, inoculated with Ag and G strains, vary significantly. This observation led us to the conclusion that disparity in symptoms evoked by these two strains may be caused by other differences in their genomes.

The most important difference between them is the presence of additional subgenomic RNA in the Ag strain, that was proven to be satellite RNA (unpublished data).

In our study we have undertaken attempts to compare the spatial model of CP protein for Polish PSV with those obtained previously for CMV, TAV, and PSV-Er [30]. Polish strains of PSV differ from other cucumoviruses analyzed before in the CP sequence. Some of these differences occur in the regions that are thought to exert an influence on mechanisms of infection, especially in two positions (R161, T191 in PSV-P and R160, T192 in PSV-G and -Ag) (Fig. 1).

Symptoms of viral infection depend highly on the inoculated host plants and result from species-specific plant–virus interaction, which in turn result from plant's resistance mechanisms and viral mechanisms to break it. In case of viruses possessing satellite RNAs, we have to take into account even more complex system, virus–satellite–plant interaction. This, additionally, makes difficult the observation of correlation between sequence and symptoms, because of the capability of satellite to modulate symptoms expression [5, 6]. Since two Polish strains, PSV-P and PSV-Ag, have satRNAs in their virions there is only one possibility to consider sequence–symptoms correlation without any ambiguity in case of PSV-G. Especially that we can observe the interfering effect of satellite when comparing symptoms caused by PSV-G and PSV-Ag, the virus strains with almost identical CPs as well as other genomes' fragments (unpublished data).

In previous experiments [10], CMV-R and CMV-Trk7 strains were found in position 193 asparagine (N) and lysine (K), respectively. The CMV-R induced strong stunting whereas CMV-Trk7 induced mosaic in *N. glutinosa* plants. Gellert et al. [30] found that this residue is located close to casein kinase II phosphorylation site. It was suggested there that charge of asparagines or other aa residues at that position might be responsible for kinase

bonding on the CPs surface. Point mutations in of CMV-Trk7 CP—K193S and K193N—caused stunting in *N. glutinosa* [10]. Because the charge of threonine (as it is in Polish PSVs) is similar as that of serine, we may assume that it may cause the same effect. We noticed that in some of the tested plants (*C. amaranticolor*, *P. sativum*, *P. vulgaris*) (Table 1) PSV-G induced systemic stunting. Lack of stunting in PSV-P and -Ag may result from satRNA's presence in their genome, or other sequence determinants as it was suggested in case of PSV-Er [30]. However, because of the very high homology of Ag and G CPs we assume that in the case of Ag strain this is rather an effect of symptom modulation by satRNA.

Presence of mosaic on those plants together with stunting may draw a remark that lysine at the position 191 in Polish strains of PSV (as it is at the corresponding position 193 in CMV-Trk7) is not necessarily required for mosaic occurrence, at least not alone or at least not in tested host plants. However, such conclusion needs huge number of tested virus strains to be reliable.

In the other potentially important positions 129 and 214 in CMV (in PSV 127/128 depending on the strain and 212/213, respectively) several wild types and mutants of CMV-M and CMV-Fny were tested on squash plants [11]. In general, from comparisons of other *Cucumovirus* strains [30] it was suggested that rigid loop containing proline 129 and/or lack of phosphorylation sites in positions 212–214 enhance symptoms of infection. In contrary, flexible  $\beta$ E- $\alpha$ EF loop (around the position 129 in CMV) might cause inhibition of infection, allowing for phosphorylation in positions 212–214, in case when such phosphorylation site exists. In Polish strains there are proline and threonine in positions corresponding to 129 and 214 in CMV. Threonine 212 could be phosphorylated, but the presence of rigid proline 127 abolishes such modification, therefore leading to the development of infection symptoms. Analyzed strains, especially PSV-G (but PSV-P in much lesser extent), elicit severe symptoms in almost all plants tested so far, which is consistent with this conclusion.

In previous study it has been demonstrated that between infectious maize CMV-Fny and non-infectious CMV-M strains there are several nt substitutions [33]. CMV-Fny at the positions 129, 162, and 168 has proline, alanine, and tyrosine residues, respectively, while CMV-M has leucine, threonine, and cysteine, respectively at those positions. Experiments with single, double, and triple mutations in CMV-M had shown that only double mutant L128P/T162 and triple mutant T129P/T162A, C168Y of CMV-M caused infections in inoculated maize plants. Therefore, it had been suggested that nt or aa sequence corresponding to positions 129 and 162 might be involved in infectious phenotype. In the case of all Polish strains, proline, arginine, and tyrosine were at the

positions 129, 162, and 168, respectively. It is, therefore, possible that lack of alanine at the position 162 is responsible for incapability for maize infection, although also other determinants may account for this phenotype in strains analyzed by our group.

Precise identification of the region in CP derived from Polish strains of PSV responsible for symptom induction or attenuation and the effect of point mutation on viral disease needs future experiments, including direct mutagenesis. Currently, we can only assume that different symptoms on tested plants infected with particular Polish strains of PSV may result from the presence of additional satRNA in case of PSV-P and -Ag, which could modulate the symptoms of disease. Therefore, in the case of those two strains it is very difficult to compare symptoms as well as to find correlations between specific amino acids residues and nature of symptoms. We suppose that satRNA may be able to attenuate the symptoms of disease, but these symptoms modulation via satRNA is dependent on the interaction between satRNA, virus, and plant host.

Modeling the N-terminus of CP was the matter of our particular interest, as the first 20 aa of CP are very rich in arginine. It had been reported, that CPs of *Bromoviridae* contains the ARM [28]. Apart of *Bromoviridae*, ARM recognizing and specifically binding RNA occurs also in proteins of other plant viruses (*Sobemovirus*, *Tombusvirus*) and in ribosomal proteins of nonplant viruses as well as in Tat and Rev proteins of HIV [28] and in bacterial antiterminators (reviewed in [34]). Functional analysis of ARM revealed its role in specific binding and encapsidation of genomic RNAs in BMV [29] and CCMV [35]. Previous research had shown that ARM present in numerous viruses from other groups can adopt frequently the helical conformation, as well as in vitro synthesized peptide, corresponding to ARM from BMV [29] and CCMV [28]. Despite the high similarity between sequences of CP from PSV genome with CP sequences from other *Bromoviridae* [28], we were not able to identify the conserved ARM characteristic for *Bromoviridae*, although the content of arginine residues in the sequence of the first 20 aa is high (33%). Despite the high similarity between sequences of CP from PSV genome with CP sequences from other *Bromoviridae*, we were not able to identify the conserved ARM characteristic for *Bromoviridae*, although the content of arginine residues in the sequence of the first 20 aa is high (33%). Predicting the structure of this region proved not doubtless also. Due to the proline residue in the motif SRRPRRGRRS (residues 10–19, reported to be responsible for RNA binding [36], helical structure of this region is highly unlikely, although in one of the generated models of PSV-Ag CP the region 10–14 had been predicted to form an  $\alpha$ -helix. The de novo modeling of this region resulted in obtaining several almost equivalent models with different

conformation and position of the disordered 19 residues from N-terminus. It cannot be excluded that the region adopts defined conformation only after binding to some other molecule. Some experiments had afforded a proof, that in the case of distantly related viruses helical conformation of ARM is not always necessary for specific RNA binding [28], still, determining the possible role of arginine-rich region of PSV CP requires mutagenesis experiments and further functional analysis.

**Acknowledgment** We thank Dr. Natasza Borodynko for the help in the infectivity tests and purification of the viruses.

## References

1. J.R. Diaz-Ruiz, J.M. Kaper, Arch. Virol. **75**, 277–281 (1983)
2. A. Twardowicz-Jakusz, H. Pospieszny, Bull. Acad. Pol. Sci. **29**, 423–427 (1983)
3. C.C. Hu, A.E. Aboul-Ata, R.A. Naidu, S.A. Ghabrial, J. Gen. Virol. **78**, 929–939 (1997)
4. L.Y. Yan, Z.Y. Xu, R. Goldbach, C. Kunrong, M. Prins, Arch. Virol. **150**, 1203–1211 (2005)
5. M. Pelczyk, A. Obrepalska-Stepulowska, H. Pospieszny, Postepy. Biochem. **52**, 212–221 (2006)
6. A.E. Simon, M.J. Roosinck, Z. Havelda, Annu. Rev. Phytopathol. **42**, 415–437 (2004)
7. M. Takeshita, M. Suzuki, Y. Takanami, Arch. Virol. **146**, 697–711 (2001)
8. P. Palukaitis, F. Garcia-Arenal, Adv. Virus Res. **62**, 241–323 (2003)
9. I.B. Kaplan, L. Zhang, P. Palukaitis, Virology **246**, 221–231 (1998)
10. D. Szilassy, K. Salanki, E. Balazs, MPMI **12**, 1105–1113 (1999)
11. S.M. Wong, S.S.C. Thio, M.H. Shintaku, P. Palukaitis, MPMI **12**, 628–632 (1999)
12. A.L. Rao, G.L. Grantham, Virology **226**, 294–305 (1996)
13. H. Pospieszny, J. Plant Protect. Res. **28**, 27–75 (1987)
14. J. Sambrook, E.F. Fritsch, T. Maniatis (eds.), *Molecular Cloning. A Laboratory Manual*, 2nd edn. (Cold Spring Harbor Laboratory Press, New York, 1989), pp. 7.43–7.45
15. S.K. Choi, J.K. Choi, W.M. Park, K.H. Ryu, J. Virol. Methods **83**, 67–73 (1999)
16. K.B. Nicholas, H.B. Nicholas Jr., D.W. Deerfield II, EMBNEW News **4**, 14 (1997)
17. J. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, Nucleic Acids Res. **25**, 4876–4882 (1997)
18. S. Kumar, K. Tamura, M. Nei, Brief. Bioinform. **5**, 150–163 (2004)
19. S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Nucleic Acids Res. **25**, 3389–3402 (1997)
20. J. Kosinski, I.A. Cymerman, M. Feder, M.A. Kurowski, J.M. Sasin, J.M. Bujnicki, Proteins **53**(Suppl 6), 369–379 (2003)
21. K.T. Simons, C. Kooperberg, E. Huang, D. Baker, J. Mol. Biol. **268**, 209–225 (1997)
22. M.A. Kurowski, J.M. Bujnicki, Nucleic Acids Res. **31**, 3305–3307 (2003)
23. A. Fiser, A. Sali, Methods Enzymol. **374**, 461–491 (2003)
24. J. Sasin, J. Bujnicki, Nucleic Acids Res. **32**(Web Server issue), W586–589 (2004)
25. N. Siew, A. Elofsson, L. Rychlewski, D. Fischer, Bioinformatics **16**, 776–785 (2000)



26. B. Wallner, A. Elofsson, *Protein Sci.* **12**, 1073–1086 (2003)
27. N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, *Proc. Natl. Acad. Sci. USA* **98**, 10037–10041 (2001)
28. R. Tan, A.D. Frankel, *Proc. Natl. Acad. Sci. USA* **92**, 5282–5286 (1995)
29. R. Sacher, P. Ahlquist, *J. Virol.* **63**, 4545–4552 (1989)
30. A. Gellert, K. Salanki, G. Naray-Shabo, E. Balazs, *J. Mol. Graph. Model.* **24**, 319–327 (2006)
31. M.R. Hajimorad, C.C. Hu, S.A. Ghabrial, *Arch. Virol.* **144**, 1587–1600 (1999)
32. T.J. Smith, E. Chase, T. Schmidt, K.L. Perry, *J. Virol.* **74**, 7578–7586 (2000)
33. K.H. Ryu, C.H. Kim, P. Palukaitis, *MBMI* **11**, 351–337 (1998)
34. C.G. Burd, G. Dreyfuss, *Science* **265**, 615–621 (1994)
35. R. Duggal, T.C. Hall, *J. Virol.* **67**, 6406–6412 (1993)
36. Y.G. Choi, A.L. Rao, *Virology* **275**, 207–217 (2000)