

RESEARCH

Open Access

# Tridimensional model structure and patterns of molecular evolution of *Pepino mosaic virus* TGBp3 protein

Beata Hasiów-Jaroszewska<sup>1\*</sup>, Anna Czerwoniec<sup>2</sup>, Henryk Pospieszny<sup>1</sup> and Santiago F. Elena<sup>3,4</sup>

## Abstract

**Background:** *Pepino mosaic virus* (PepMV) is considered one of the most dangerous pathogens infecting tomatoes worldwide. The virus is highly diverse and four distinct genotypes, as well as inter-strain recombinants, have already been described. The isolates display a wide range on symptoms on infected plant species, ranging from mild mosaic to severe necrosis. However, little is known about the mechanisms and pattern of PepMV molecular evolution and about the role of individual proteins in host-pathogen interactions.

**Methods:** The nucleotide sequences of the triple gene block 3 (TGB3) from PepMV isolates varying in symptomatology and geographic origin have been analyzed. The modes and patterns of molecular evolution of the TGBp3 protein were investigated by evaluating the selective constraints to which particular amino acid residues have been subjected during the course of diversification. The tridimensional structure of TGBp3 protein has been modeled *de novo* using the Rosetta algorithm. The correlation between symptoms development and location of specific amino acids residues was analyzed.

**Results:** The results have shown that TGBp3 has been evolving mainly under the action of purifying selection operating on several amino acid sites, thus highlighting its functional role during PepMV infection. Interestingly, amino acid 67, which has been previously shown to be a necrosis determinant, was found to be under positive selection.

**Conclusions:** Identification of diverse selection events in TGBp3 will help unraveling its biological functions and is essential to an understanding of the evolutionary constraints exerted on the *Potexvirus* genome. The estimated tridimensional structure of TGBp3 will serve as a platform for further sequence, structural and function analysis and will stimulate new experimental advances.

**Keywords:** molecular evolution, PepMV, protein modeling, selective constraints, TGBp3, virus evolution

## Background

*Pepino mosaic virus* (PepMV) belongs to *Potexvirus* genus within the *Flexiviridae* family and it is a well-known pathogen infecting tomato plants worldwide. PepMV possesses a single-stranded, positive-sense RNA genome of approximately 6.4 kb, flanked by 5' and 3' untranslated regions (UTRs) with a 5' cap and a 3' poly(A) tail. PepMV contains five conserved open reading frames (ORFs). ORF1 encodes a 164 kDa replication-related protein with

three functional domains: an N-terminal mRNA capping enzyme, a central RNA helicase, and a C-terminal RNA-dependent RNA polymerase (RdRp). ORF2 through ORF4 encode movement proteins (TGBp1, TGBp2 and TGBp3) of 26, 14, and 9 kDa, respectively. These three ORFs overlap in the genome and, hence, are called the triple gene block. The 25 kDa CP is encoded by ORF5 [1-3]. The virus is highly diverse; four different genotypes (CH2, LP, EU, and LP) and a large number of inter-strain recombinants have been described [4,5]. PepMV isolates differ both in host range as well in the symptoms induced on susceptible plants species. Recently, it was proved that one single mutation in TGBp3 protein converts a mild

\* Correspondence: B.Hasiow@iorpib.poznan.pl

<sup>1</sup>Institute of Plant Protection-National Research Institute, ul. Wł. Węgorka 20, 60-318 Poznań, Poland

Full list of author information is available at the end of the article

pathotype of CH2 genotype into a necrotic one. This is the only association reported between particular amino acid sites and symptomatology [6]. However, our preliminary results with isolates from the EU genotype showed that we are dealing with a universal mechanism of variability. By performing site-directed mutagenesis of infectious clones of the EU genotype, we confirmed that amino acid 67 is responsible for inducing necrosis on tomato, irrespective of the strain (unpublished results). This finding suggests that TGBp3 structure is important for virulence, since substitution of certain amino acids results in changes in the infectivity of the virus. As the tridimensional structure of TGBp3 is not yet available, *de novo* modeling represents a suitable and convenient alternative for analyzing the relationship between amino acids present at certain positions, structure and virulence. This is the approach undertaken in this study.

A second aim of this study is to verify the nature of the selective pressures operating on TGBp3. Nucleotides substitutions can be classified into two categories on the basis of their effects on the protein amino acid composition: synonymous and nonsynonymous replacements. In general synonymous substitutions accumulate at rate  $d_S$  per synonymous site. It is generally accepted that synonymous substitutions accumulate neutrally because they have no effect on the amino acids composition and thus shall not affect the folding and function of proteins. In contrast, nonsynonymous substitutions accumulate at rate  $d_N$  per synonymous site. Since these substitutions involve amino acid replacements, they must be subjected to selection. The average ratio of these two substitutions rates  $\omega = d_N/d_S$  across a gene, provides information on whether it has been fixing amino acid replacements in a neutral fashion ( $\omega = 1$ ), amino acids changes have been removed by the action of purifying selection ( $\omega < 1$ ), or changes have been fixed by positive evolution ( $\omega > 1$ ) [7]. However, averaging  $\omega$  across the entire coding sequence of a gene is a gross approximation since highly conserved and hypervariable sites may coexist. It is more sensitive to allow the  $\omega$  rates ratio to vary among sites and to take the presence of individual sites at which  $\omega > 1$  as an evidence for positive selection [7]. In this study, sequence analysis of the PepMV TGB3 was conducted and  $\omega$  rates ratio were calculated. Additionally, a computational analysis of TGB3 sequences to investigate the extent of recombination events was performed. Taken together, in this study we present results of computational analyses evaluating the patterns of molecular evolution and their effect on protein structure and virulence.

## Methods

### Virus isolates and phylogenetic analysis

Eighteen sequences of PepMV TGB3 representing isolates from different genotypes and geographic regions

were retrieved from GenBank and used for further analysis (Table 1). For some isolates detailed biological information was also available. Sequences were translated into amino acids using BioEdit [8]. Multiple protein sequence alignments were obtained using MUSCLE [9] and then protein-coding nucleotide sequence alignments were constructed based on their corresponding protein sequence alignments using PAL2NAL[10].

It is well known that recombination affects the estimation of  $\omega$  and interferes with phylogenetic reconstruction [11]. Therefore the PepMV sequences were initially analyzed for evidence of recombination events using the RDP package [12]. The sequences were analyzed using the following methods: 3 Seq, Chimaera, Lard, Bootscan, RDP, Genecovn, and MaxChi. Only recombination breakpoints supported by more than three methods were considered as valid. With this stringent criterion, no significant recombination signals were identified in the analyzed sequences. The best-fitting model of nucleotide substitution was investigated using the MODELTEST implemented in MEGA5; Kimura 2-parameters model (K2P) was chosen. Maximum-likelihood phylogenetic tree was constructed using MEGA5 and K2P model of nucleotide substitution [13]. Reliability of the obtained tree was evaluated using the bootstrap method based on 1000 pseudoreplicates. Both nucleotide and amino acids sequences were scrutinized to find particular positions distinguishing mild and severe isolates.

**Table 1 Isolates used in study**

Isolate	Origin	Genotype	Symptoms on <i>S. lycopersicum</i>	Accession Numbers
1806	Belgium	EU	mild	FJ457098
DB1	Netherlands	EU	necrosis	FJ940224
SW	Poland	EU	mild	EF408822
UK	UK	EU	-	AF340024
SP13	Spain	EU	-	AF484251
PMU0618	Spain	EU	-	FJ263318
H	Hungary	EU	mild (yellow mosaic)	AM491606
No	Poland	EU	mild	FJ612602
LP2001	Peru	LP	asymptomatic	AJ606361
US3	USA	EU	-	AY508411.1
Ch1	Chile	CH1	-	DQ000984.1
P22	Poland	CH2	mild	HQ650560
Cl308	Spain	CH2	-	GU130097
PMU0848	Spain	CH2	-	FJ263365
P19	Poland	CH2	necrosis	HQ650559
Ch2	Chile	CH2	-	DQ000985
Amo7	Spain	CH2	-	GU130090
US2	USA	recombinant CH1/CH2	-	AY509927

### Analysis of selective pressure

Our analysis of selective pressures was based on evaluating  $\omega$  for each codon in the nucleotide sequence alignment. The single-likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL) and internal branches fixed-effects likelihood (IFEL) methods were used to evaluate  $\omega$  per codon [14]. Both SLAC and FEL methods used using the default significance level of  $p = 0.1$ ; a Bayes factor of 40 was used as selection threshold for REL. These analyses were performed using the HyPhy package [15] as implemented in the DATAMONKEY webserver [16]. Nucleotide substitutions were modeled according to the K2P scheme. If a gene evolves under purifying selection most of the time but is occasionally subject to episodes of positive selection, a comparison between two distant related sequences is unlikely to yield a  $\omega$  ratio significantly  $> 1$ . To specifically test whether phylogenetic branches with necrotic isolates underwent positive selection events, we performed two independent analyses of branch-specific codon models. First, the M2 branch site model allowing  $\omega$  to vary among sites in the protein and across branches on the tree and aim to detect positive selection affecting a few sites along particular lineages [17] was evaluated using the program CODEML from the PAML version 4.4 package [18]. Second, we applied the sliding-window analysis implemented in SWAPSC [19]. SWAPSC uses a statistically optimized window size to detect selective constraints in specific codon regions of the alignment at a particular branch of phylogenetic tree. The method estimates the null distribution of  $d_S$  and  $d_N$  from simulated sequence alignments. A statistically optimal window size is then estimated that makes the detection of adaptive evolution independent of the window size. Four codons was the estimated optimal window size. Simulated sequences were generated with the program EVOLVER from the PAML package [18] with parameters estimated from the true sequence alignments after running M0 codon-based model in CODEML.

### Identification of functionally important amino acid residues

To identify potentially functional and structural amino acids residues we used the Bayesian methods implemented in the ConSurf server [20]. Functionally important residues, e.g. involved in ligand binding and protein-protein interactions, are often evolutionarily conserved and are most likely to be solvent-accessible, whereas conserved residues within the protein core most probably have an important structural role in maintaining the protein properly folded [20].

### Secondary structure prediction and mutation mapping

Secondary structure prediction and tertiary fold-recognition (FR) were performed using the GENESILICO meta-

server gateway [21]. Secondary structure was predicted using PSIPRED [22], PROFSEC [23], PROF [24], SABLE [25], and JNET [26]. FR methods did not report any good template structure for homology modeling procedure. Hence, *de novo* methods for structure modeling were applied.

### *De novo* structural modeling of TGBp3

To obtain a tridimensional structural model of the TGB3 protein we used TGBp3 sequences of P19, P22 and DB1 isolates. P19 and P22 represent necrotic and mild pathotypes of the CH2 genotype, respectively. DB1 isolates were described as necrotic isolates belonging to the EU genotype. The ROSETTA [27] algorithm was used for *de novo* modeling of TGBp3. Hundreds of thousands of decoys were generated and clustered to identify the best low-energy conformations. Selection of models based on the average energy clusters, size, density and visual inspection of final structures was performed. The quality of the final models was evaluated using the neural network based algorithm implemented in the PROQ server [28,29] and the MetaMQAPII meta-server [30]. In the first case, the quality is quantified by the location in the plane formed by the two indexes *LGscore* (i.e., the  $-\log$  of a  $p$ -value) and *MaxSub* (ranging 0 - 1). Depending on the specific values of these indexes, the model can be qualified as: correct if *LGscore*  $> 1.5$  and *MaxSub*  $> 0.1$ , as good if *LGscore*  $> 3$  and *MaxSub*  $> 0.5$ , and as very good if *LGscore*  $> 5$  and *MaxSub*  $> 0.8$ . Mapping of the electrostatic potential on protein surfaces was calculated with APBS (Adaptive Poisson-Boltzmann Solver) [31]. This procedure was performed to reveal potential electrostatic changes on the surface on the protein that may influence interaction between other proteins and elements in local cell environment.

## Results

### Analysis of TGBp3

The analysis of TGBp3 sequences revealed striking differences between particular isolates. TGBp3 consists of 246-258 nt and 82-86 amino acids, respectively, depending on the isolate considered. The protein of the Peruvian isolate LP-2001 from *Lycopersicon peruvianum* has two extra amino acids in the C-terminal region (86) than the characteristic 82-84 of other isolates. It is unclear whether this polymorphism in TGBp3 length has any affect on symptoms. The nucleotide identity between the 18 sequences studied ranged from 82% to 98.7% and from 77.3% to 98.7% at the amino acid level. The average transition/transversion rates ratio was high, 3.9, as observed for other plant RNA viruses [32].

The phylogenetic tree shown in Figure 1 illustrates the presence of two major groups containing isolates from

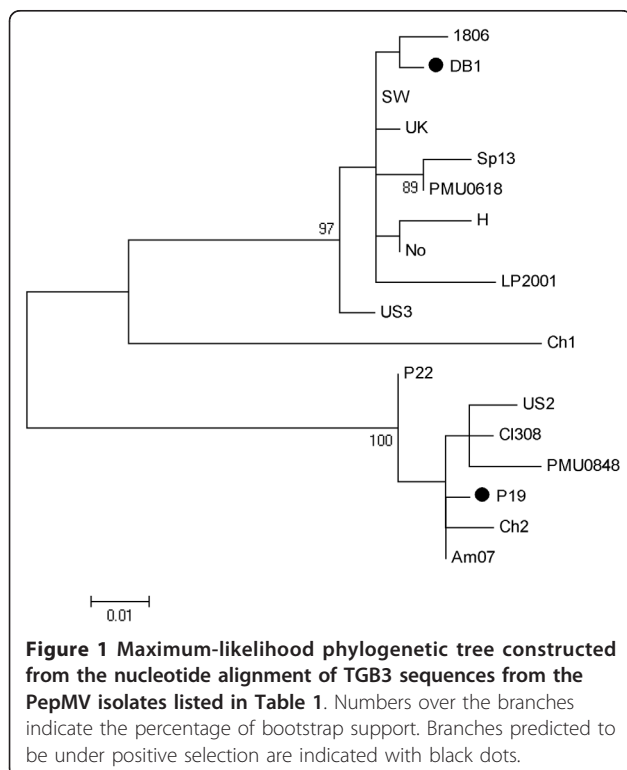
CH2 and EU genotypes. CH1 isolate form a separate branch. Both SLAC and REL method found no evidence of positively selected codons, but found three negatively selected ones: 50, 63 and 65 with  $\omega = 0.36$ . By contrast, the FEL method predicted codon 37 as to be under positive selection, plus 15 other codons to be under purifying selection. Both CODEML and SWAPSC analyses revealed low  $\omega$  value (0.35), thus suggesting a strong action of purifying selection in the evolution of TGB3 among the isolates studied. As mentioned above,  $\omega < 1$  is indicative of purifying selection removing deleterious nonsynonymous mutations. The whole TGBp3 has undergone purifying selection in the majority of branches of the phylogenetic tree. However, branch-site model implemented in CODEML showed positive selection acting in branches driving to necrotic isolates (Figure 1). In seven branches negative selection operated on region encompassing nucleotides 187-198 (amino acids 63-66), that includes the highly conserved PEVL motif. Those residues were also classified as functionally or structurally important by the Bayesian methods implemented in ConSurf. However, amino acid 67 (nucleotides 199-201) was predicted to be under positive selection, despite being surrounded by amino acids under negative selection. In this position, a K was found in the majority of PepMV isolates and only isolates DB1 (EU genotype) and P19 (CH2 genotype) contained E at this position, a very radical change replacing a basic

residue by an acid one. Both isolates have been described as necrotic.

### Structural modeling of TGBp3

Secondary structure prediction with PSIPRED [22] revealed the existence of three  $\alpha$ -helices in regions 2-13, 17-36 and 62-68 and two  $\beta$ -strands in regions 45-50 and 53-57. Consistently, the modeled tridimensional structure preserved these structures. Position 67, which undergoes mutation K to E, is located in the third  $\alpha$ -helix, without affecting the helix stability. A transmembrane domain/helix was predicted for the region encompassing amino acids 10 to 35 (Figure 2).

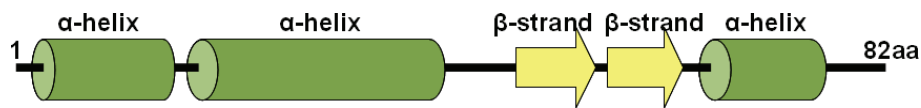
A three-dimensional model for the TGBp3 from the P19 necrotic isolate was obtained using the ROSETTA algorithm. The model was evaluated as "extremely good" according to the *LG score* (5.751) and "correct" according to the *MaxSub* index (0.320). The METAMQAPII method predicted that the overall GDT\_TS score (Global Distance Test Total Score - a measure of similarity between two protein structures with identical amino acid sequences, but different tertiary structures) of the final model with respect to the native (currently unknown) structure is 59.45 (GDT\_TS ranges between 1 and 100) and the model's root mean square deviation to the true structure is around 2.6. This indicates that the model can be used to make functional inferences at the level of individual residues. Overall, the inferred model for the TGBp3 (82 amino acids) consists of two antiparallel  $\beta$ -sheets flanked by three  $\alpha$ -helices which exactly corresponds to those predicted by PSIPRED (Figure 3). Based on the structure of P19 isolate TGBp3, we modeled two other proteins: one from the mild isolate P22 and one from the necrotic isolate DB1. These two additional proteins were chosen to evaluate the effect of mutations conferring the necrotic phenotype on the protein structure. The amino acid sequence identity of TGBp3 between isolates P19 and DB1 was 85% and between P19 and P22 was as high as 99%. This level of amino acid identities allowed us to use homology-modeling methods [33] to build tridimensional models of both proteins. We particularly analyzed mutation K to E at residue 67. This mutation occurs rarely in PepMV homologous sequences. It is predicted to be in the surface of the folded protein and locally changing potential on the surface of helical region with mutation from negative to positive (Figure 4). This small change may influence on local interactions with environment components as well as the process of protein folding into native shape, but this needs further investigation.



### Discussion

Many plant virus genera encode a triple gene block (TGB), an especially evolutionarily conserved gene

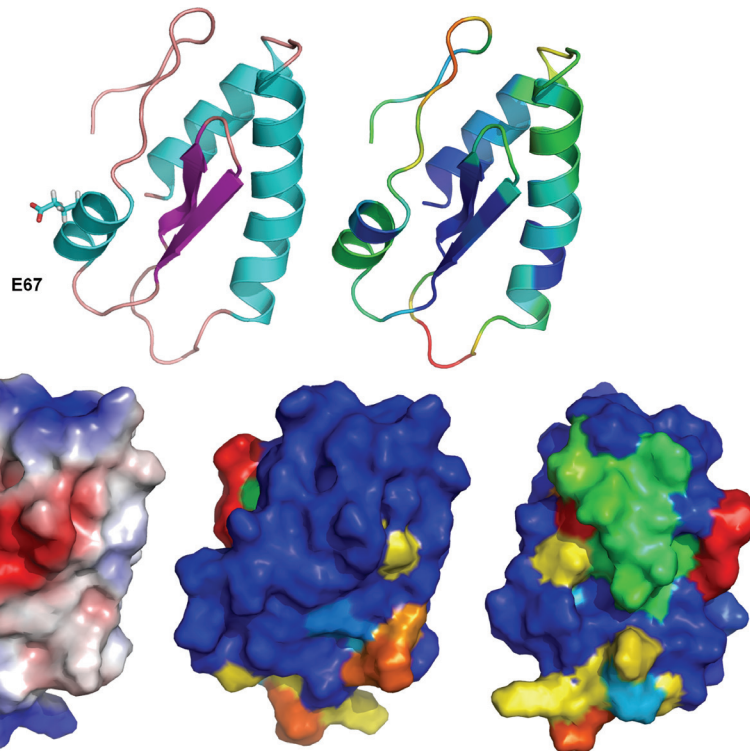




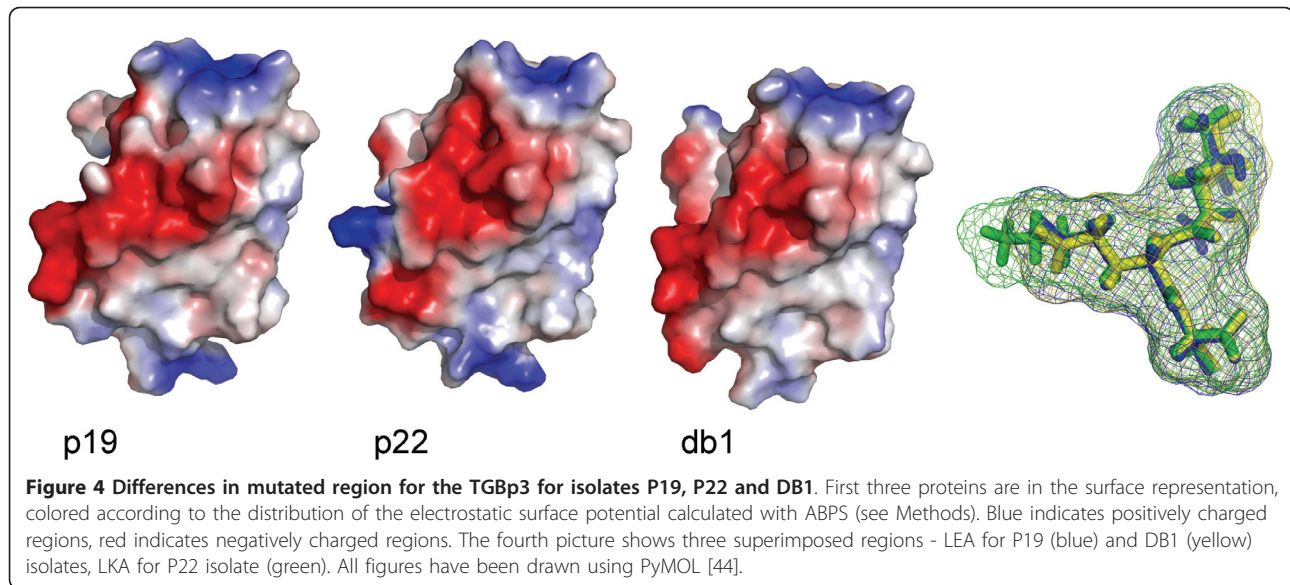
**Figure 2** Consensus results of the secondary structure prediction.  $\alpha$ -helices are shown as green tubes and  $\beta$ -strands as yellow arrows.

module involved in the cell-to-cell and long-distance movement of viruses. The TGB-based transport system exploits the coordinated action of three polypeptides to deliver viral genomes into plasmodesmata and to accomplish virus entry into neighboring cells [34,35]. TGB-encoded proteins are referred to as TGBp1, TGBp2 and TGBp3, according to the positions of their cistrons [35]. All three proteins are essential for virus movement. TGBp1 was widely studied and besides its role in movement, it has been shown that generally functions as an RNA silencing suppressor in members of the genus *Potexvirus* [36]. At the same time, little is known about additional function of TGBp2 and TGBp3. In agreement with sequence analysis and *in vitro* studies

predicting that TGBp2 and TGBp3 are integral membrane proteins [37] cell fractionation of plant tissues expressing these proteins demonstrates predominant association of both proteins with the P1 and P30 membranous fractions as well as with the cell wall [38,39]. Understanding the molecular evolutionary biology of the various proteins expressed by viral genomes and their functions is a prerequisite for the control of virus propagation and the elaboration of efficient and durable antiviral strategies. In our previous study we have shown that TGBp3 is involved in host-pathogen interactions during PepMV infection. Experiments with PepMV TGBp3 mutants revealed that one single mutation K67E was required for converting a mild pathotype into a



**Figure 3** Structural model of TGBp3. The first panel shows a model in the ribbon representation colored according to the secondary structure elements. The second panel shows a model colored according to the predicted local deviation from the real structure, as calculated by METAMQAPII. 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 third panel shows a protein in the surface representation, colored according to the distribution of the electrostatic surface potential calculated with ABPS (see Methods). Blue indicates positively charged regions, red indicates negatively charged regions. The fourth and fifth panels show models in the surface representation, colored according to sequence similarity, calculated from multiple sequence alignments of PepMV sequences using ConSurf. All four models are in the same orientation, except the last one, which has been rotated horizontally. All figures have been drawn using PyMOL [44].



necrotic one [6]. Mutant viruses of mild PepMV strain induced necrosis on *Datura innoxia* and *Solanum lycopersicum*. Symptoms of viral infection strongly depended on the inoculated host plant and result from species-specific host-pathogen interactions.

Secondary and tertiary TGBp3 structure predictions showed that the protein consists of three  $\alpha$ -helices and two  $\beta$ -strands. The secondary structure prediction placed amino acid 67 in one of the  $\alpha$ -helices. Our tridimensional structural predictions revealed that the region encompassing amino acid 67 in isolates P19, P22 and DB1 is located on the surface of the protein and thus mutations in this region, specially when the physical properties dramatically change as it is the case of mutation K67E, must have a strong impact on the ability of TGBp3 to interact with other protein. It is well known that in many cases amino acids can be replaced without impairing protein function, even if these are of quite different physico-chemical characteristics. However, the change of a positively charged K by a negatively charged E may change the local property of the protein surface, jeopardizing its ability to establish the correct interactions with other viral proteins or cell components. Supporting the existence of such functionality in this surface region, we have identified nucleotide sites 187-198, spanning amino acids 63-66 that were under the action of negative selection in seven branches of the phylogenetic tree describing the evolutionary history of PepMV. In particular, the PEVL motif is highly conserved in all isolates analyzed.

In general, functionally essential protein parts are negatively selected for (conserved), while other parts can be positively selected for. The  $\omega$  mean value obtained for the TGBp3 cistron strongly suggests a

predominant action of purifying selection. In good agreement with this average value, three different approaches detected strong signal of purifying selection for particular codons. Nevertheless, most codons are evolutionarily neutral. However, perhaps the most interesting result from the analyses of selective constraints is that amino acid K67 has been identified as under positive selection in the branch leading to necrotic isolates, suggesting that necrosis may be an adaptive trait. Since it is close to the region of amino acids 63-66, which is (i) under negative selection (and hence likely has an important function) and (ii) predicted to be in the surface of the folded protein, we can speculate that it may be involved in the formation of protein-protein complex that determine the development of symptoms. The results obtained with ConSurf indicated that amino acids 61-66 played functional and structural roles in the protein. A protein function is however, the results of the functional and structural communication between sites and, therefore, the ability of a given site to change depends on the interactions it must establish with other residues of the molecule. Mutations at either nearby sites (like K67E), or functionally related distant sites in the structure, will change the selective constraints [40]. Functional sites, like binding domains, are less prone to amino acid changes than less important protein regions. Furthermore, some classes of proteins evolve faster than others [41].

The analysis of the TGB1 gene in PepMV populations clearly provides a mechanism for its rapid evolution and adaptation to the ever-changing environments [42]. In the light of what is so far known about PepMV evolutionary dynamics [43] it seems that TGBp3 evolves mostly by the action of purifying selection operating

over several sites, highlighting its functional role during PepMV infection. Gómez et al. [43] showed only one amino acid under purifying selection in the TGBp3 of CH2 genotype also using the methods implemented in the DATAMONKEY server. We have analyzed TGBp3 sequences representing different genotypes and we used more sophisticated methods to establish selection pressure acting on TGBp3. We were able to identify more codons under action of purifying selection, moreover amino acid 67 was predicted to be under positive selection. It seems that the particular pathotypes achieved an advantage over others especially in Europe. Recently, more aggressive pathotypes (causing necrosis or severe yellowing) of CH2 genotype have become dominant in Poland. This shift in the Polish PepMV population reveals a dynamic interplay between the different PepMV genotypes and their host. Necrotic isolates, however, were described in both the EU and CH2 genotypes, and it seems that the selective pressures act in the direction of increasing the virulence of isolates, less temperature dependent which cause significant losses in quality and quantity of yield. Preliminary data (Hasiów-Jaroszewska, unpublished results) have shown that necrotic isolates acquired the ability to infect a larger number of *Solanum tuberosum* varieties and of causing more severe symptoms in shorter time. Probably, the key feature of these isolates are faster replication or accumulation. It seems that molecular evolution is leading to higher variability among CH2 genotype in comparison to others genotypes. It has been suggested that the CH2 genotype has a biological advantage over the EU genotype, as it seems to spread more rapidly within a crop [5]. Recent study on the evolutionary dynamics of the PepMV population in Spain using RT-quantitative PCR analyses in inoculated tomato plants showed that a CH2 isolate (PS5) accumulated more rapidly and to higher viral loads than an EU isolate (Sp13) [43]. The TGBp3 is responsible for virus movement and further research will be performed to shed the light between particular mutation in TGBp3 and virus fitness.

The present *in silico* study opens new research avenues for researchers interested in experimentally exploring the *in vivo* interactions between TGBp3 and host factors. Moreover, we report the first tridimensional structure of TGBp3, obtained with *de novo* folding methods followed by careful accuracy assessments. This model may serve as a platform for further sequence, structural and function analysis and will stimulate new experiment advances.

## Conclusions

The obtained results suggested that TGBp3 has been evolving mainly under the action of purifying selection operating on several amino acid sites however amino

acid 67, which has been previously shown to be a necrosis determinant, was found to be under positive selection. The estimated tridimensional structure of TGBp3 will serve as a platform for further sequence, structural and function analysis and will stimulate new experimental advances.

## Acknowledgements

This study was funded by the Polish Ministry of Science and Higher Education grants 0067/P01/2010/70 (AC) and N N310 163 438 and IP2010 012470 Iuventus Plus (to BHJ). The study was also supported by Foundation for Polish Science. Work in Valencia was supported by the Spanish MICINN grant BFU2009-06993 (to SFE) and by EMBO Short Term Fellowship ASTF424-2010 (to BHJ).

## Author details

<sup>1</sup>Institute of Plant Protection-National Research Institute, ul. Wł. Węgorka 20, 60-318 Poznań, Poland. <sup>2</sup>Bioinformatics Laboratory, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Umultowska 89, PL-61-614 Poznań, Poland. <sup>3</sup>Instituto de Biología Molecular y Celular de Plantas, CSIC-UPV, 46022 València, Spain. <sup>4</sup>The Santa Fe Institute, Santa Fe, NM87501, USA.

## Authors' contributions

BHJ and SFE participated in the design of the study. BHJ and AC performed analyses and drafted the manuscript. SFE and HP participated in drafting the final version of manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

Received: 11 May 2011 Accepted: 24 June 2011 Published: 24 June 2011

## References

1. Maroon-Lango CJ, Guaragna MA, Jordan RL, Hammond J, Bandla M, Marquardt SK: Two unique US isolates of *Pepino mosaic virus* from a limited source of pooled tomato tissue are distinct from a third (European-like) US isolate. *Arch Virol* 2005, **150**:1187-1201.
2. Ling KS: Molecular characterization of two *Pepino mosaic virus* variants from imported tomato seed reveals high levels of sequence identity between Chilean and US isolates. *Virus Genes* 2007, **34**:1-8.
3. Hasiów B, Borodynko N, Pospieszny H: Complete genomic RNA sequence of the Polish *Pepino mosaic virus* isolate belonging to the US2 strain. *Virus Genes* 2008, **36**:1-8.
4. Hasiów-Jaroszewska B, Borodynko N, Jackowiak P, Figlerowicz M, Pospieszny H: *Pepino mosaic virus* - a pathogen of tomato crops in Poland: biology, evolution and diagnostics. *J Plant Protect Res* 2010, **50**:477-483.
5. Hanssen I, Thomma B: *Pepino mosaic virus*: a successful pathogen that rapidly evolved from emerging to endemic in tomato crops. *Mol Plant Pathol* 2010, **11**:179-189.
6. Hasiów-Jaroszewska B, Borodynko N, Jackowiak P, Figlerowicz M, Pospieszny H: A single mutation in TGB3 converts mild pathotype of *Pepino mosaic virus* into necrotic one. *Virus Res* 2011, **159**:57-61.
7. Yang Z, Bielawski JP: Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* 2000, **15**:496-503.
8. Hall TA: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 1999, **41**:95-98.
9. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res* 2004, **32**:1792-1797.
10. Suyama M, Torrents D, Bork P: PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucl Acids Res* 2006, **34**:609-612.
11. Silva MC, Edwards SV: Structure and evolution of a new avian MHC Class II B gene in a sub-antarctic seabird, the thin-billed prion (Procellariiformes: *Pachyptila belcheri*). *J Mol Evol* 2009, **68**:279-291.



12. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P: **RDP3: a flexible and fast computer program for analyzing recombination.** *Bioinformatics* 2010, **26**:2462-2463.
13. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011.
14. Kosakovsky Pond SL, Frost SDW: **Not so different after all: A comparison of methods for detecting amino acid sites under selection.** *Mol Biol Evol* 2005, **22**:1208-1222.
15. Kosakovsky Pond S, Frost SDW, Muse SV: **HyPhy: hypothesis testing using phylogenies.** *Bioinformatics* 2005, **21**:676-679.
16. Kosakovsky Pond SL, Frost SDW: **DATAMONKEY: rapid detection of selective pressure on individual sites of codon alignments.** *Bioinformatics* 2005, **21**:2531-2533.
17. Yang Z, Nielsen R, Goldman N, Pedersen AMK: **Codon-substitution models for heterogeneous selection pressure at amino acid sites.** *Genetics* 2000, **155**:431-449.
18. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 2000, **13**:555-556.
19. Fares MA: **SWAPSC: sliding-window analysis procedure to detect selective constraints.** *Bioinformatics* 2004, **20**:2867-2868.
20. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-T N: **ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids.** *Nucl Acids Res* 2010, **38**:529-533.
21. Kurowski MA, Bujnicki JM: **GeneSilico protein structure prediction meta-server.** *Nucl Acids Res* 2003, **31**:3305-3307.
22. McGuffin LJ, Bryson K, Jones DT: **The PSIPRED protein structure prediction server.** *Bioinformatics* 2000, **16**:404-405.
23. Rost B, Yachdav G, Liu J: **The PredictProtein server.** *Nucleic Acids Res* 2004, **32**:321-326.
24. Ouali M, King RD: **Cascaded multiple classifiers for secondary structure prediction.** *Protein Sci* 2000, **9**:1162-1176.
25. Adamczak R, Porollo A, Meller J: **Accurate prediction of solvent accessibility using neural networks-based regression.** *Proteins* 2004, **56**:753-767.
26. Cuff JA, Barton GJ: **Application of multiple sequence alignment profiles to improve protein secondary structure prediction.** *Proteins* 2000, **40**:502-511.
27. Simons KT, Kooperberg C, Huang E, Baker D: **Assembly of protein tertiary structure from fragments with similar local sequences using simulated annealing and bayesian scoring function.** *J Mol Biol* 1997, **268**:209-225.
28. Siew N, Elofsson A, Rychlewski L, Fischer D: **MaxSub: An automated measure to assess the quality of protein structure predictions.** *Bioinformatics* 2000, **16**:776-785.
29. Wallner B, Elofsson A: **Can correct protein models be identified.** *Protein Sci* 2003, **12**:1073-1086.
30. Pawlowski M, Gajda MJ, Matlak R, Bujnicki JM: **MetaMQAP: a meta-server for the quality assessment of protein models.** *BMC Bioinformatics* 2008, **9**:403.
31. Baker NA, Sept D, Holst MJ, McCammon JA: **Electrostatics of cellular components: Application to microtubules and the ribosome.** *Proc Natl Acad Sci USA* 2001, **98**:10037-10041.
32. Tomas N, Elena SF: **The rate and spectrum of spontaneous mutations in a plant RNA virus.** *Genetics* 2010, **185**:983-989.
33. Kosinski J, Cymerman IA, Feder M, Kurowski MA, Sasin JM, Bujnicki JM: **A 'Frankenstein's monster' approach to comparative modeling: merging the finest fragments of fold-recognition models and iterative model refinement aided by 3D structure evaluation.** *Proteins* 2003, **53**:369-79.
34. Morozov SY, Solovyev AG: **Triple gene block: modular design of a multifunctional machine for plant virus movement.** *J Gen Virol* 2003, **84**:1351-1366.
35. Beck DL, Guilford PJ, Voot DM, Andersen MT, Forster RLS: **Triple gene block proteins of *White clover mosaic potexvirus* are required for transport.** *Virology* 1991, **83**:695-702.
36. Senshu H, Ozeki J, Komatsu K, Hashimoto M, Hatada K, Aoyama M, Kagiwada S, Yamaji Y, Namba S: **Variability in the level of RNA silencing suppression caused by triple gene block protein 1 (TGBp1) from various potexviruses during infection.** *J Gen Virol* 2009, **90**:1014-24.
37. Morozov SY, Miroshnichenko NA, Solovyev AG, Zelenina DA, Fedorkin ON, Lukashva LI, Grachev SA, Chernov BK: **In vitro membrane binding of the translation products of the carlavirus 7-kDa protein genes.** *Virology* 1991, **183**:782-785.
38. Cowan GH, Lioliopoulou F, Ziegler A, Torrance L: **Subcellular localisation, protein interactions, and RNA binding of potato mop-top virus triple gene block proteins.** *Virology* 2002, **298**:106-115.
39. Gorshkova EN, Erokhina TN, Stroganova TA, Yelina NE, Zamyatin AA Jr, Kalinina NO, Schiemann J, Solovyev AG, Morozov SY: **Immunodetection and fluorescent microscopy of transgenically expressed hordeivirus TGBp3 movement protein reveals its association with endoplasmic reticulum elements in close proximity to plasmodesmata.** *J Gen Virol* 2003, **84**:985-994.
40. Tourasse NJ, Li WH: **Selective constraints, amino acid composition, and the rate of protein evolution.** *Mol Biol Evol* 2000, **17**:656-664.
41. Farfan M, Minana-Galbis D, Carmenn Fuste M, Loren G: **Divergent evolution and purifying selection of the *flaA* gene sequences in *Aeromas*.** *Biol Direct* 2009, **4**:23-39.
42. Hasiów-Jaroszewska B, Jackowiak P, Borodynko N, Figlerowicz M, Pospieszny H: **Quasispecies nature of *Pepino mosaic virus* and its evolutionary dynamics.** *Virus Genes* 2010, **41**:260-267.
43. Gómez P, Sempere RN, Elena SF, Aranda MA: **Mixed infections of *Pepino mosaic virus* strains modulate the evolutionary dynamics of this emergent virus.** *J Virol* 2009, **83**:12378-12387.
44. DeLano WL: **The PyMOL User's Manual.** DeLano Scientific, San Carlos, CA, USA; 2002.

doi:10.1186/1743-422X-8-318

**Cite this article as:** Hasiów-Jaroszewska et al.: Tridimensional model structure and patterns of molecular evolution of *Pepino mosaic virus* TGBp3 protein. *Virology Journal* 2011 **8**:318.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

