Vol. 10(2), pp. 15-31, March, 2019
DOI: 10.5897/JBSA2019.0111
Article Number: 06DF3CF60644
ISSN 2141-2464

http://www.academicjournals.org/JBSA

ACADEMIC JOURNALS

Journal of Bioinformatics and Sequence Analysis

Full Length Research Paper

Computational characterization of providence virus non-structural proteins: Evolutionary and functional implications

Nakayinga Ritah

Department of Biological Sciences, Faculty of Science, Kyambogo University, Kampala, Uganda.
Department of Biochemistry, Microbiology and Biotechnology, Faculty of Science, Rhodes University, Grahams town, South Africa.

Received 9 January, 2019; Accepted 27 March, 2019.

Providence virus is the only member of the family Carmotetraviridae and carries a positive single stranded RNA genome that encodes three open reading frames. The smallest open reading frame encodes the structural proteins. The largest open reading frame encodes a large putative protein, p130. The second overlapping open reading frame encodes two non-structural proteins; p40, a proposed accessory protein and p104, the replicase, containing the RdRp domain. Till date, p130 and p40 are not associated with any related open reading frames in the databases. The purpose of this study is to identify sequences within these non-structural proteins with potential roles in replication and evolution using computational tools. Our results revealed that p130 has a putative arginine-rich sequence which lies in the disordered region also found in the *Umbravirus*, *Groundnut rosette virus* p27. Analysis of the p40 revealed a sequence with a coiled-coil conformation and surface-exposed characteristics comparable to the interaction domain of *Tombusvirus*, *Tomato bushy stunt virus* p33 accessory protein. The hypothetical two transmembrane helix topology of PrV p104 oriented the putative localization signal at the N-terminus, the same way the localization signal of *Tomato bushy stunt virus* p92 is oriented. This study concluded that Providence virus non-structural proteins are structurally related to *Tombusvirus* and *Umbravirus* accessory proteins and contain sequences with predicted functions in replication. Findings from this study have led us to propose a co-evolutionary event between an insect and plant virus resulting in a hybrid virus with the potential to infect and replicate in both host plant and animal systems.

Key words: Providence virus, non-structural proteins, p40, p130, sequence comparison, replication, evolution.

INTRODUCTION

Providence virus (PrV) represents the only member within the family Carmotetraviridae and carries a carmo-(RdRp) motifs conserved among members of Tombusviridae and Umbraviridae (Walter et al., 2010).

E-mail: rnakayinga@kyu.ac.ug. Tel: +256703369516.

The positive single stranded RNA (+ssRNA) virus was initially discovered in a persistently infected *Helicoverpa zea* (*H. zea*) midgut cell line and is the only tetravirus able to replicate in tissue culture (Pringle et al., 2003; Jiwaji et al., 2016). Tetraviruses are insect viruses classifying into three families: *Alphatetraviridae*, *Permutotetraviridae*, and Carmotetraviridae, according to the nucleotide sequence of the viral replicase (Dorrington et al., 2011). Tetraviruses are +ssRNA ,non-enveloped with a characteristic *T* = 4 capsid symmetry and limited host range with order Lepidoptera and Chiroptera (Moore, 1991; Bawden et al., 1999; Pringle et al., 2003; Kemenesi et al., 2016). The monopartite genome of PrV differs from other tetraviruses in that it encodes three open reading frames (ORF) instead of the typical two ORFs (Walter et al., 2010). The putative viral replicase ORF (p104) and viral capsid precursor ORF (p81) are conserved among all tetraviruses (Walter et al., 2010). The presence of a read through stop type 1 signal, UAGCAACUA, within the replicase results in the production of the accessory protein, p40 and full-length protein, p104 characterized with RdRp motifs, required for the establishment of infection (Walter et al., 2010). The third and largest ORF, p130, overlaps the replicase gene and is unique to PrV. The protein consists of a putative 2A-like processing site (PrV-$2A_1$) whose activity is predicted to produce two translation products of 17 kDa and 113 kDa and is functional in *in vitro* studies (Walter et al., 2010; Luke et al., 2008).

The translational control system for the expression of PrV replicase resembles that typically observed in tombusviruses. For instance, the expression of the *Tomato bushy stunt virus* (TBSV) genome results in the replicase (p92) and accessory protein (p33) via the ribosomal readthrough amber termination signal (Scholthof et al., 1995). Within p33 is a short p33:p33/p92 interaction domain important for mediating protein-protein between itself and p92 (Panavas et al., 2005; Rajendran and Nagy, 2004, 2006). The RNA binding sequence, RPRRRP, present in p33 is important for binding genomic RNA (Rajendran and Nagy, 2003). The two transmembrane domains (TMD), TMD1, TMD2 and peroximal targeting signals are responsible for membrane anchorage and localization of p33 and p92 onto the surface of the peroxisomal membrane, the site for assembly of the replication complex and viral RNA synthesis (McCartney et al., 2005).

The Umbraviridae genome comprises four ORFs: The ORF1 (accessory protein), ORF2 (RdRp), ORF3 (RNA chaperon protein) and ORF4 (movement protein) (Ryabov et al., 2011). The *Groundnut rosette virus* (GRV) ORF3 proteins possess three functions that include RNA chaperone activities, protection of viral RNA against plant defensive RNA silencing systems and mediating long distance movement through the phloem (Taliansky et al., 2003). Unique to umbraviruses is the absence of a capsid gene and therefore lack the ability to produce conventional virus particles in infected plants (Taliansky and Robinson, 2003; Taliansky et al., 2003). Within the host, umbraviruses use ribonucleoproteins, made from complexes of GRV ORF3 protein and genomic RNA, as alternatives to classic capsid proteins to shuttle viral RNA via long distance movement through the phloem and to establish systemic viral infection (Taliansky et al., 2003; Kim et al., 2007).

So far, aspects of tetravirus replication have been limited to studies by subcellular localization with viral RNA and replication proteins. These studies have shown that the replicase of *Helicoverpa armigera* stunt virus, *Alphatetraviridae*, associates with membranes derived from endosomes while PrV replicase associates with membrane vesicles from the Golgi apparatus and secretory pathway (Walter, 2008; Short et al., 2010, 2013).

As a first step towards understanding the replication biology of PrV, this study seeks to identify sequences within PrV non-structural proteins with potential functions in replication using computational tools. Till date, no identifiable ORFs or known peptide homologues have been reported in p130 and p40 and their evolutionary origins remain unknown. The aim of this study is to begin to functionally characterize PrV non-structural proteins using computational tools with regards to their potential roles in replication and shed light on their evolutionary origins.

MATERIALS AND METHODS

Prediction of potentially functional sequences in PrV p130

The BLAST search engine at https://www.ncbi.nlm.nih.gov was employed to search for protein homologues of PrV p130. The putative RNA binding region was identified by visual inspection of a sequence that resembles the amino acid residues of the RNA binding arginine-rich motif of GRV ORF3 (Taliansky et al., 1996). The p130 sequence was submitted to IUPRED program (http://iupred.enzim.hu/) for identification of putative disordered or unstructured sequences. The long disorder prediction parameter was selected for this analysis. The putative loop structures were identified by Predict Protein server site at http://www.predictprotein.org/.

Sequence comparison of the putative interaction sequence in PrV p40

The search for conserved domains in PrV p40 was performed by BLAST search engine at https://www.ncbi.nlm.nih.gov. The Protscale program hosted by the Expert Protein Analysis System (ExPASy) (http://web.expasy.org/protscale) was used to identify putative surface-exposed and hydrophilic sequences in PrV p40. The Kyte and Doolittle and Hopp and Woods scales were selected along with a window size of 5 for both methods. Potential coiled-coil sequence in p40 was identified by the Predict Protein server at http://www.predictprotein.org/. The sequence was submitted to the server site and analyzed using amino acid window sizes 14, 21, 28.

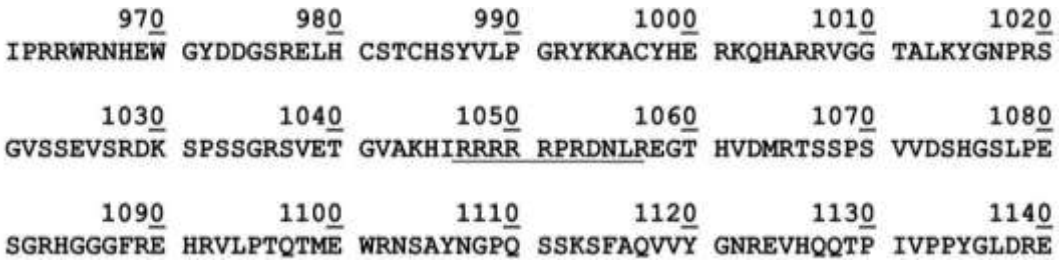


Figure 1. Identification of a putative RNA-binding sequence in PrV p130. Partial sequence of p130 showing the putative RNA binding sequence indicated with a solid line below the amino acid sequence. The positions of the amino acids are depicted above the protein sequence.

Potential RNA binding sequence in PrV p104

The PrV arginine-rich sequence, RRRRYA, at amino acids position 476 to 481, shares amino acids R, Y and A with Tombusvirus p33 arginine/proline rich motif, was analyzed for its surface-exposed and hydrophilic properties using the Protscale program (http://web.expasy.org/protscale) using a window size of 5. The scales Kyte and Doolittle and Hopp and Woods were selected for this function.

Prediction of transmembrane helix and topology of PrV p104

Putative transmembrane helix was generated using TMpred (http://www.ch.embnet.org/software/TMPRED_form.html) and Protscale Kyte and Doolittle scale at http://web.expasy.org/protscale, based on window size 19. This window size type displays hydrophobic membrane-spanning sequences with strong signals above the threshold value of 1.5. The topology of PrV p104 was derived at by comparison analysis with the membrane topology of TBSV p92 (Scholthof et al., 1995), given that both viral proteins share evolutionary and functional similarities (Walter et al., 2010).

Putative subcellular localization signals of PrV non-structural proteins

PrV p104 and p130 sequences were queried using the Signal IP 4.0 program at http://www.cbs.dtu.dk/services/SignalP/, in order to predict putative signal sequences with subcellular targeting functions.

RESULTS

Identification of the putative RNA binding sequence and disordered region in PrV p130

The BLAST search engine did not identify any protein homologues or conserved domains in PrV p130 amino acid sequence. The evolutionary relatedness between PrV, tombusviruses and umbraviruses (Walter et al., 2010) led to the search for amino acid sequences shared between PrV p130 and the non-structural proteins of tombusviruses and umbraviruses. An arginine-rich sequence, RRRRRPRDNLR, was found at amino acids 1047 to 1057 of PrV p130 (Figure 1). This sequence shares arginine residues with the arginine-rich sequence, RPRRRAGRSGGMDPR, at amino acid positions 108 – 122, of Umbravirus GRV ORF3 that encodes p27 (GRV p27). This protein is an RNA binding protein (Taliansky et al., 2003).

The putative disordered or unstructured region within PrV p130 was detected by the IUPRED programme. The returned graphical output displayed a region that spans amino acids 517 - 1220 with scores above the threshold value of 0.5 that is considered statistically significant for being a putative disordered region (Figure 2A). The arginine-rich sequence translated in the PrV p130 ORF lies within the disordered region (Figure 2A), a feature that might have an important biological function. For comparison, the GRV p27, an RNA binding protein, was analysed by the IUPRED programme and the results displayed almost the entire protein sequence with a score above the threshold value 0.5 (Figure 2B). The arginine-rich sequence of GRV p27 (Kim et al., 2007) resides within the predicted disordered region (Figure 2B). In both proteins, the arginine-rich sequences were located within the disordered region.

The predicted secondary structure type generated by Predict Protein program of the amino acids 517 - 1220 of the disordered region of p130 consisted mainly loop structures (Supplementary Figure S1). Similarly, the Predict Protein-generated results for GRV p27 showed that the protein sequence was entirely covered with loop structures (Supplementary Figure S2)

Prediction of a protein-protein interaction sequence in PrV p40

Since the BLAST search engine did not indicate any functional domains in p40 amino acid sequence, the evolutionary relatedness between PrV and tombusvirus (Walter et al., 2010) led to search for similar sequences in p40 and non-structural proteins of tombusviruses. Sequence comparisons were carried out to identify sequences in p40 with similar characteristics to tombusvirus, TBSV, interaction domain (ID). The TBSV p33 ID (Panavas et al., 2005; Rajendran and Nagy, 2004; 2006)

Figure 2a. Identification of potential disordered regions. A. PrV p130. **B.** GRV p27. The plots created by IUPRED program show scores above or below the threshold value 0.5 (indicated as a line). Regions above 0.5 are disordered while regions below 0.5 are ordered. The line below the cut-off line represents position of arginine-rich sequence. Y axis represents scores while X axis represents position of the amino acids in the sequence.

Figure 2b. Identification of potential disordered regions. A. PrV p130. B. GRV p27. The plots created by IUPRED program show scores above or below the threshold value 0.5 (indicated as a line). Regions above 0.5 are disordered while regions below 0.5 are ordered. The line below the cut-off line represents position of arginine-rich sequence. Y axis represents scores while X axis represents position of the amino acids in the sequence.

was examined for its surface-exposed characteristics using Kyte and Doolittle scale and the results were used to identify a similar sequence in PrV p40 with comparable characteristics. The TBSV p33 ID located at the C-terminal amino acids, 241-282, displayed peaks with the highest signal showing a negative score of -3.0 of being surface-exposed (Figure 3A). Examination of PrV p40 using a similar scale displayed peaks with the highest signal showing a negative score of -2.5 of being surface-exposed in the region that spans amino acids 292-332 located at the C-terminal region of p40 (Figure 3B).

The Predict Protein program consolidates a range of methods and databases for predicting protein structural features. The NCOILS method detects sequences with the potential to adopt coiled-coil conformations. The p40 sequence was submitted to the server site and generated text results showing amino acids, 317 to 330, with a 30% probability of adopting a coiled-coil conformation. The putative sequence with a coiled-coil conformation was detected at the window size of 14 and not 21 or 28 (Supplementary Figure S3). The sequence was identified at the C-terminal region of p40 and coincides with the surface-exposed region predicted by ProtScale program.

Potentially functional RNA binding sequence in PrV p104

Within PrV p104 and downstream of the read through stop codon is an arginine-rich sequence, RRRRYA, at

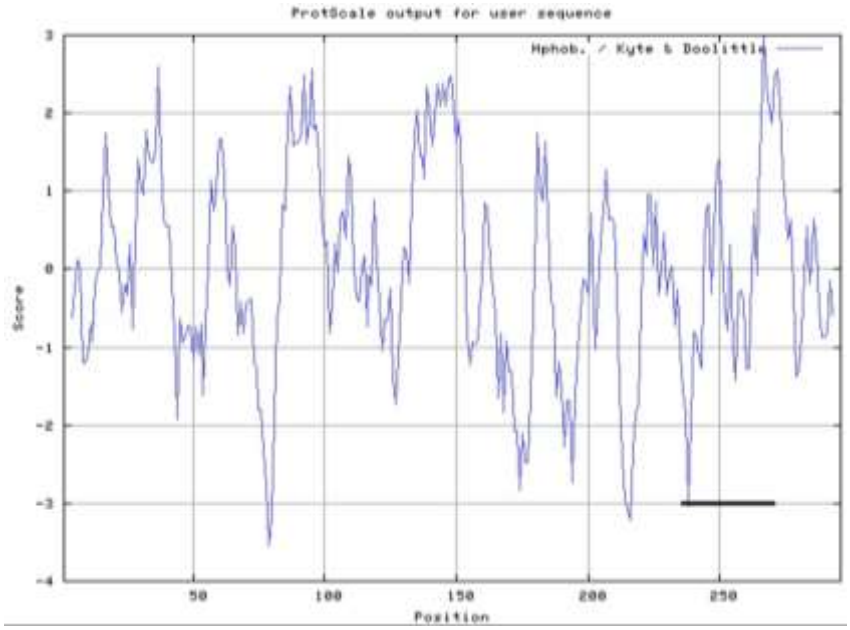


Figure 3a. Identification of a sequence with potential protein-protein interaction function in PrV p40 by comparison with surface-exposed properties of the TBSV p33 interaction domain. (A) The surface-exposed region of TBSV p33 interaction domain is underlined with a black line.

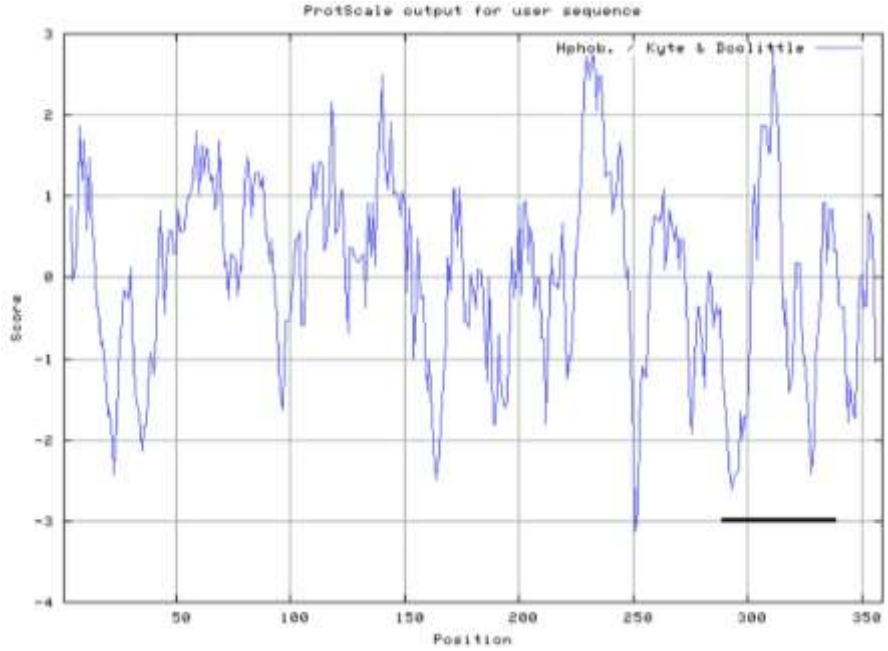


Figure 3b. The surface-exposed sequence of the putative PrV p40 interaction sequence is underlined with a black line. The Kyte and Doolittle plot was generated using a window size of 5 and surface-exposed regions are below 0. The Y axis represents scores while the X axis represents the position of the amino acid sequence.

amino acids 476 to 481. A related sequence with the sequence, RPRRRPYA, is an RNA binding motif (RBR 1) within TBSV p33 and constitutes surface-exposed and hydrophilic properties (Rajendran and Nagy, 2003). A ProtScale analysis of the PrV p104 arginine-rich motif returned a strong negative peak with a score of -4.2 of

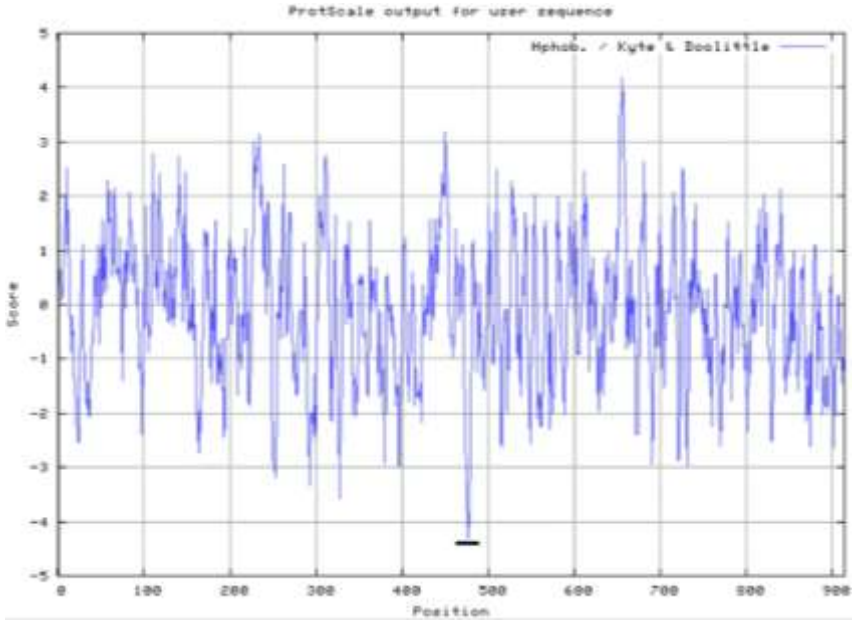


Figure 4a. Identification of surface-exposed and hydrophilic characteristics of PrV p104 arginine-rich sequence. A. The surface-exposed region is depicted by a short line below the peak. Regions below 0 are potentially surface-exposed. The Kyte and Doolittle plot was generated using a window size of 5.

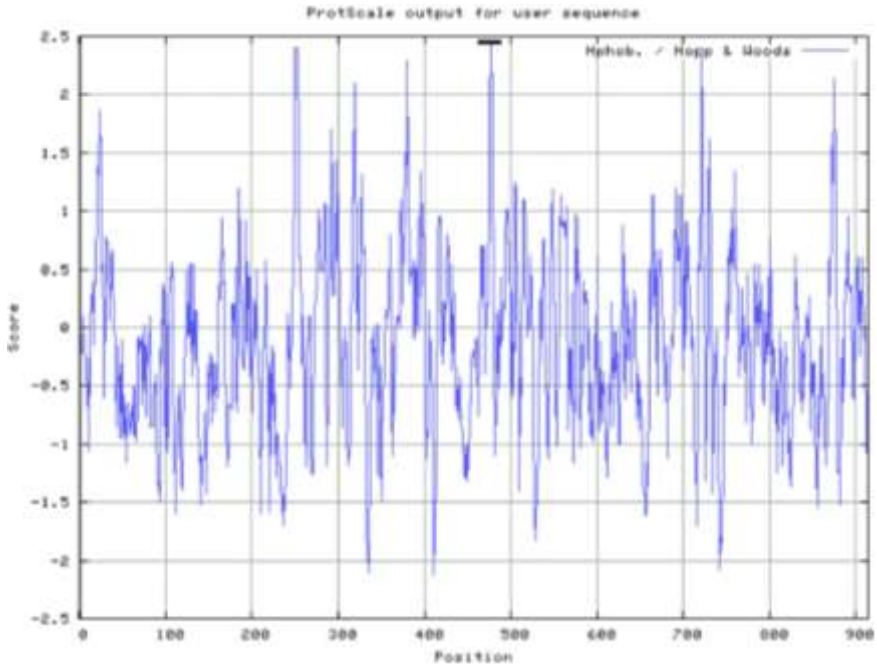


Figure 4b. B. The hydrophilic region is depicted by a short line above the peak. The Hopp and Woods plot was generated using a window size of 5 and putative hydrophilic regions are above 0. Y axis represents scores while X axis represents position of the amino acid sequence.

being potentially surface-exposed and a score of 2.45 of being potentially hydrophilic using the Kyte and Doolittle (Figure 4A) and Hopp and Woods scales (Figure 4B).

Putative transmembrane helix in PrV p104

Two programs, TMpred and Kyte and Doolittle were used

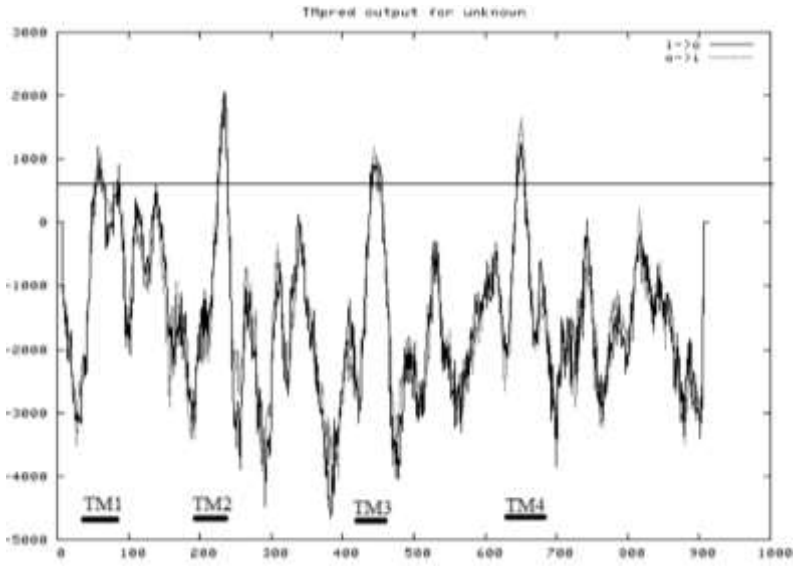


Figure 5a. Identification of potential transmembrane helix in PrV p104. A. Hydrophobicity plot generated by TMpred with scores above 500, indicated with a thick line, are considered statistically significant for transmembrane helix prediction.

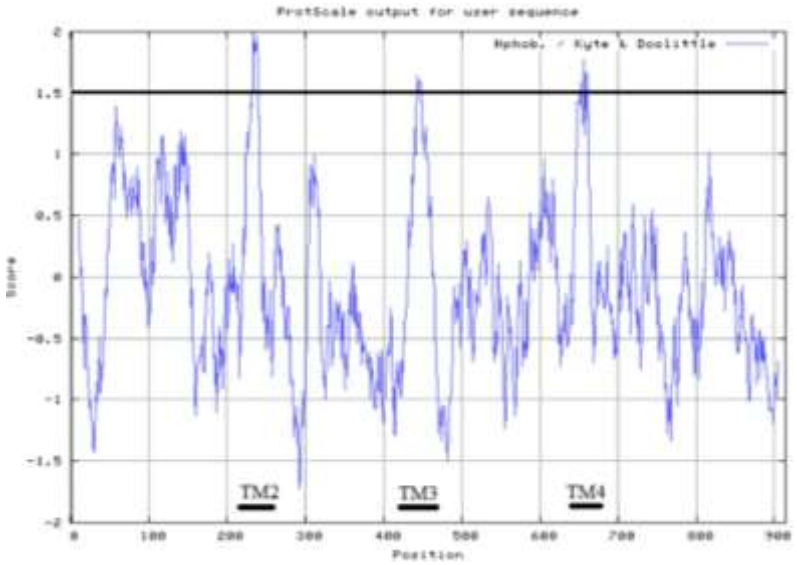


Figure 5b. Kyte-Doolittle hydropathy plot using a window scale of 19. The line at a score of 1.5 represents the threshold value for predicting transmembrane helix. The transmembrane helix; TM1 (generated by TMpred), TM2, TM3 and TM4 (generated by TMpred and Kyte-Doolittle) are depicted as black bars below the peaks.

to analyze putative transmembrane (TM) helix in PrV p104. The returned output by TMpred displayed four distinct peaks with scores above 500 and is considered statistically significant for TM helix prediction. The peaks were displayed at amino acid positions 52 - 72 (TM1), 225 - 247 (TM2), 437 - 456 (TM3) and 647 - 661 (TM4) with scores of 1101, 1949, 1213 and 1638 respectively (Figure 5A). To confirm TMpred interpretations, the Kyte and Doolittle scale based on the window size of 19 revealed three strong peaks that were well conserved and occupied the same position as those revealed by TMpred. These were 225 - 247 (TM2), 433 - 456 (TM3) and 647 - 661 (TM4) with scores 1.989, 1.642 and 1.779 respectively (Figure 5B). The putative TM1 was not

detected by Kyte and Doolittle (Figure 5B) and this could possibly be a result of the differences between algorithms used. Overall, the predicted transmembrane helix in p104 would potentially traverse the lipid membrane four or three times.

Prediction of PrV p104 topology and membrane targeting signals

Based on the transmembrane domains predicted by TMpred and Protscale programs, three possible models for the topology of p104 have been considered. The two transmembrane topology model with helix TM1 (52 - 72, identified by TMpred) and TM2 (225 - 247, identified by TMpred, Kyte and Doolittle) (Figure 6A), the three transmembrane topology model based on both algorithms (Supplementry Figure S4) and the four transmembrane topology model based on TMpred results (Supplementry Figure S5). Both the four and three transmembrane models were eliminated because the RdRp domains by convention are supposed to be exposed to the cytoplasmic side of the membrane. The transmembrane model of choice was the double transmembrane model since the two putative N-terminal helix upstream of the read through stop codon (Walter et al., 2010) would traverse the hypothetical lipid membrane twice. This would ensure that the C-terminal region of p104 which also consists of the RdRp domain (505 - 768), read through stop signal (Walter et al., 2010), putative PrV p40 ID, RBR 1 were positioned within the cytoplasm (Figure 6A). The orientation of the PrV p104 two transmembrane topology model resembles that of TBSV p92 that consists of two N-terminus transmembrane helix, 83 - 98 (TM1) and 132 - 154 (TM2), which position the RdRp domain, read through stop codon, interaction domain, RNA binding domains (RBR 1, RBR 2, RBR 3) and peroxisome targeting signal (1- 81) within the cytoplasm (McCartney et al., 2005) (Figure 6B). Also, the identification of the potential membrane targeting signal using SignalP in p104 was unsuccessful however the double transmembrane PrV p104 topology model positions the region that spans amino acids 1- 52 in the cytoplasm (Figure 6A), the same way as in TBSV p92 topology.

DISCUSSION

The aim of this study was to predict potentially functional and evolutionary conserved sequences and structures in PrV non-structural proteins. In this study, we revisit computational analysis of PrV non-structural proteins and we showed that the C-terminal sequence of p130 has an arginine-rich sequence that lie within the disordered region and covers 703 amino acid residues, most of which are mainly loop structures. Sequence comparison of p40 revealed a putative stretch at the C-terminus with surface-exposed characteristics comparable to hydrophilic characteristics of the TBSV p33 interaction domain. Secondary structure prediction analysis identified a coiled-coiled conformation within the surface-exposed C-terminus sequence of p40. The predicted transmembrane topology of p104 consists of two transmembrane helix and the structural orientation locates the putative subcellular signal at the N-terminus. Lastly, the arginine-rich sequence, RRRRYA, is present in p104 and shares amino acid residues with the RNA binding arginine/proline rich motif of Tombusvirus p33.

Results based on sequence comparison revealed a putative disordered region abundant with loop structures and an arginine-rich sequence at the C-terminal region of PrV p130. Similar characteristics were predicted for Umbraviral GRV p27 peptide. The GRV p27 is an RNA chaperone that binds genomic RNA via the arginine-rich sequence in a cooperative manner (Taliansky et al., 2003). A general view regarding RNA chaperones is that they possess long disordered domains, stretching up to 900 amino acid residues, and allow versatile conformations that interact and loosen the misfolded RNA structure without use of ATP (Herschlag, 1995; Tompa and Csermely, 2004). The proteins of both viruses share arginine-rich amino acid sequences that lie within disordered regions consisting mainly amino acids that adopt loop conformations. These unstructured flexible, arginine-rich regions may indicate a conserved structural feature that has important roles in interactions between partner molecules. This raises the possibility that PrV p130 may function as an RNA chaperone.

Study findings based on the surface-exposed prediction algorithm revealed that C-terminal amino acid residues exposed on the surface of p40 have signal scores similar to those belonging to the tombusvirus, TBSV p33 interaction domain, also located at the C-terminus and essential for mediating self-interaction between p33 molecules and with p92 replicase (Panavas et al., 2005; Rajendran and Nagy, 2004, 2006). Within the PrV p40 surface-exposed region lies a sequence with the potential of adopting an alpha helical supercoil conformation that is characteristic of protein-protein interaction functions (Lupas and Gruber, 2005; Wang et al., 2012). The similarity in location, surface-exposed characteristics with tombusvirus, TBSV p33 interaction domain and structural conformations with protein interaction properties, all together may indicate a probable protein-protein interaction sequence in PrV p40, essential for PrV RNA replication (Walter, 2008; Short et al., 2013).

A comparative topological structural analysis based on TBSV p92 transmembrane topology (Scholthof et al., 1995) showed that the two transmembrane topology model for PrV p104 was likely since it positioned all functional motifs and domains within the cytosolic side of the membrane. The putative transmembrane helix, TM1 (52 - 72) and TM2 (225 – 247) located at the N-terminus

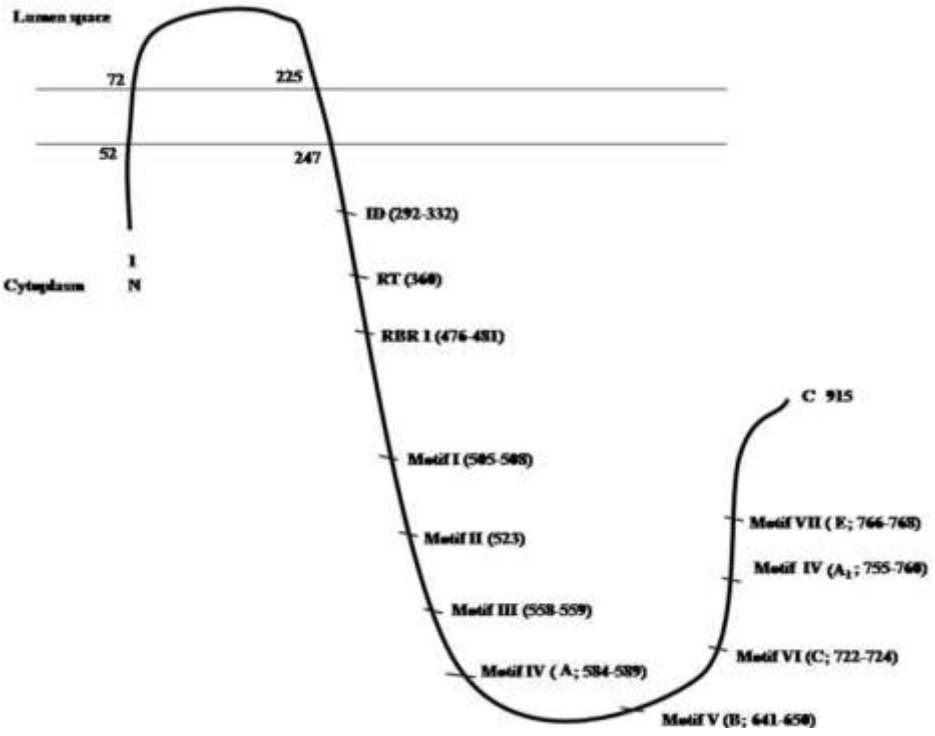


Figure 6a. The two transmembrane topology model of PrV p104 and TBSV p92. A. PrV p104.

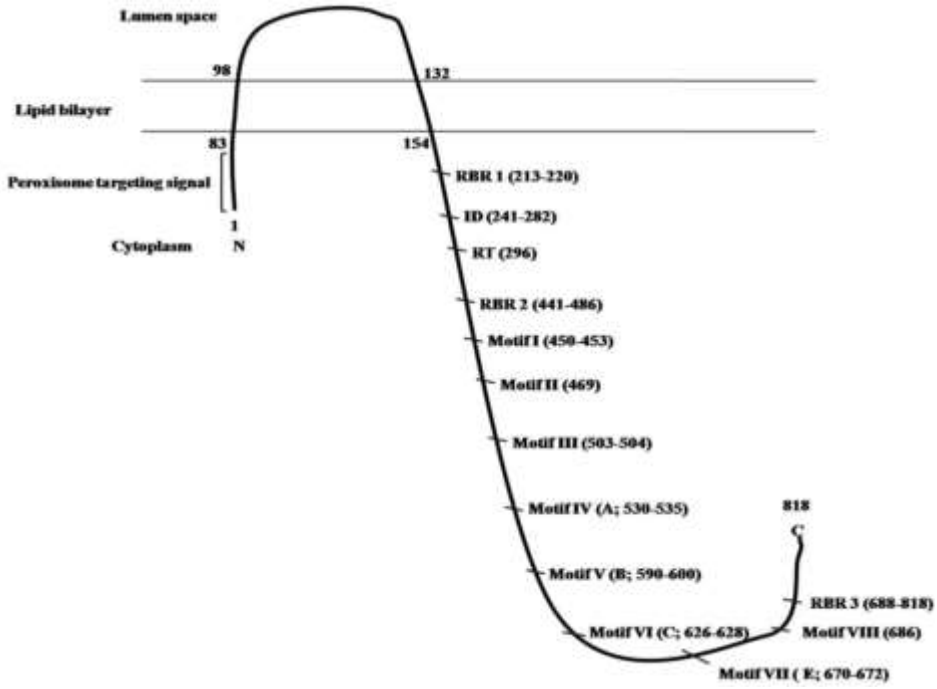


Figure 6b. TBSV p92. A hypothetical lipid bilayer is depicted as an open space showing the lumen space and cytoplasm on either side. The amino acid sequence is depicted as a black line traversing the lipid bilayer twice. The amino acid stretch of each transmembrane domain is annotated at either side of the lipid bilayer. Membrane targeting signals are indicated on the left of the protein sequence. The RdRp motifs, ID, RBR1 and RBR 2 (RNA binding region), RT (read through stop codon) are labeled on the sequence, each with its amino acid position relative to PrV p104.

of p40 would traverse the lipid membrane twice, just like in TBSV p92, and provide membrane anchorage, a prediction supported by the association of PrV replication proteins with detergent resistant membranes *in vivo* (Short and Dorrington, 2012). The two transmembrane helix found in PrV consist of 20 and 22 amino acids respectively and are comparable to those belonging to plant protein, TBSV p92, 15 and 22 amino acids respectively. The orientation of both proteins across the bilayer is identical with an amino acid stretch located at the extracellular side and two amino acid segments at the intracellular side of the bilayer. The first 52 amino acids of p104 were hypothetically located at the cytosolic side of the membrane, identical to the orientation of TBSV p92 transmembrane topology (Scholthof et al., 1995), and may contain subcellular membrane targeting signals that direct the viral replication complex (VRC) to membranes derived from the Golgi apparatus and/or secretory pathway (Short et al., 2013).

The arginine-rich sequence, RRRRYA, found at the C-terminal region of PrV p104 shares amino acids R, Y, A, with TBSV p33 arginine/proline-rich motif important for binding RNA (Rajendran and Nagy, 2003). Sequence comparison revealed that the p104 arginine-rich sequence lies within the hydrophilic and surface-exposed region, identical to those described for the RPR motif in TBSV p33 (Rajendran and Nagy, 2003). These predictions suggest that the arginine-rich sequence within the protein is potentially exposed to the cytoplasmic environment.

The current knowledge of PrV evolution is based on morphological, structural and comparative genomics because a reverse genetic system to rescue these viruses is unavailable. The virus shares the host range, capsid morphology and genome organisation with other tetraviruses (Dorrington et al., 2011), all evidence of an evolutionary relationship with insect viruses. On the contrary, PrV replicase shares a read through stop signal, expression strategy and replicase sequence similarity with plant viruses of the carmo-like super group II, suggesting a conserved protein expression strategy with plant viruses (Walter et al., 2010). The prediction of disordered-loop structures, arginine-rich sequences and protein-protein interaction sequences in p130 and p40 was also found in plant viruses from the carmo-like super group II; it further supported an evolutionary structural relationship with carmo-like plant viruses. The data led us to propose that PrV non-structural proteins are of a plant virus origin yet the structural similarities with insect viruses make PrV a potentially chimeric virus that may have evolved as a result either by convergent evolution from an insect and plant virus ancestor or is the evolutionary result of an insect and plant virus co-infection *in vivo*. The assumption is that PrV may have acquired plant viral proteins via horizontal transfer between insects and plants during insect feeding sessions on plants. These acquired proteins may have important contributions to viral movement, protection and replication in the plant host. All together, PrV presents as a virus that has the capacity to infect and replicate in both plant and insect systems thereby expanding its host range.

Conclusion

In this study, computational investigations into PrV non-structural proteins revealed sequences with potential functions in replication and evolution previously unknown. A putative arginine-rich sequence lies within the disordered region with amino acids that adopt loop confirmations in PrV p130. A region within p40 shares surface-exposed and hydrophilic characteristics with tombusvirus, TBSV p33 interaction domain and adopts a coiled-coil conformation. The two transmembrane helix topology model of p104 orients the probable membrane localization signal within the cytoplasmic side of the plasma membrane, same as TBSV p92 transmembrane helix topology. The identification of sequences within PrV highlights important roles in replication; however, experimental evidence is needed to develop a replication model that will allow better understanding of PrV replication biology. The similarity in structural and sequence characteristics directly supports the evolutionary relatedness of PrV non-structural proteins with sequences belonging to plant viruses. This might have important implications on PrV host range with a possibility of extending to plant systems.

CONFLICT OF INTERESTS

The author has not declared any conflict of interests.

REFERENCES

Bawden AL, Gordon KHJ, Hanzlik TN (1999). The specificity of *Helicoverpaarmigera* stunt virus infectivity. Journal of Inverterbrate Pathology 74(2):156-163.

Dorrington RA, Gorbalenya AE, Gordon KHJ, Lauber C, Ward VK (2011). *Tetraviridae*. In Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses, Edited by King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ. San Diego: Elsevier Academic Press pp. 1091-1102.

Herschlag D (1995). RNA chaperones and the RNA folding problem. Journal Biological Chemistry 270(36):20871-20874.

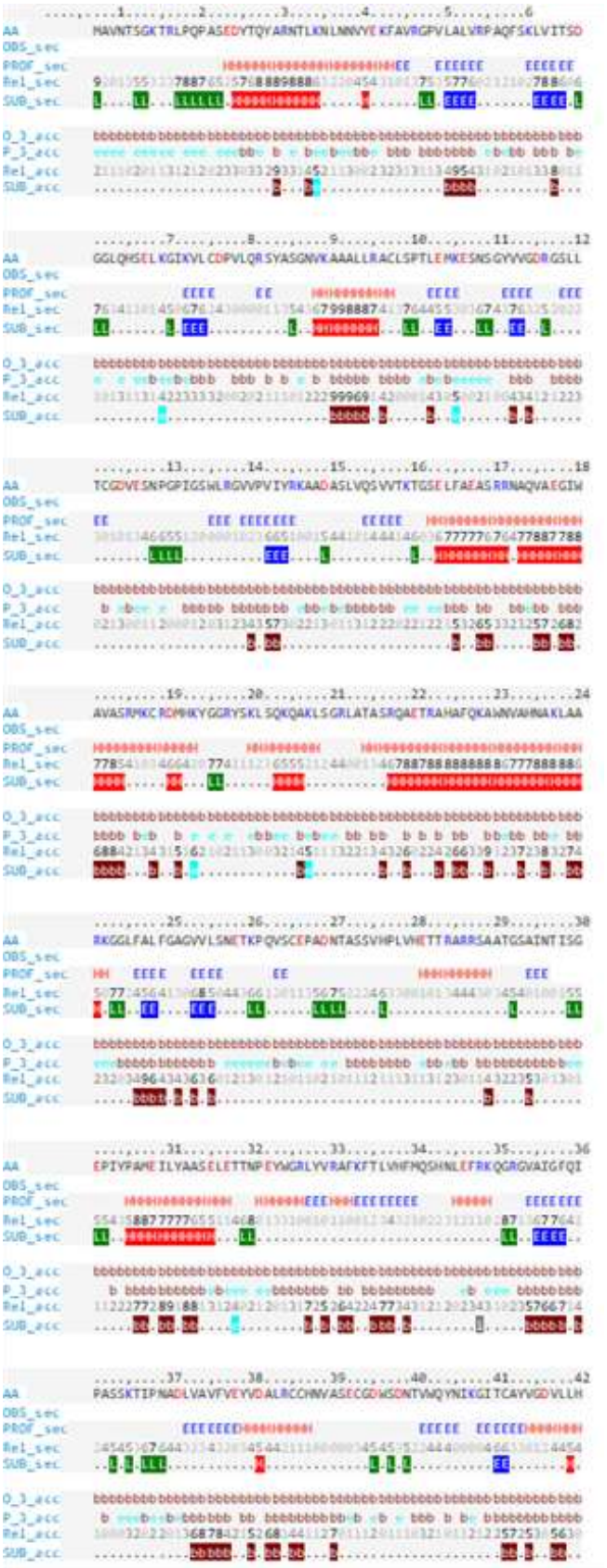
Jiwaji M, Short JR, Dorrington RA (2016). Expanding the host range of small insect RNA viruses: Providence virus (Carmotetraviridae) infects and replicates in a human tissue culture cell line. Journal General Virology 97(10):2763-2768.

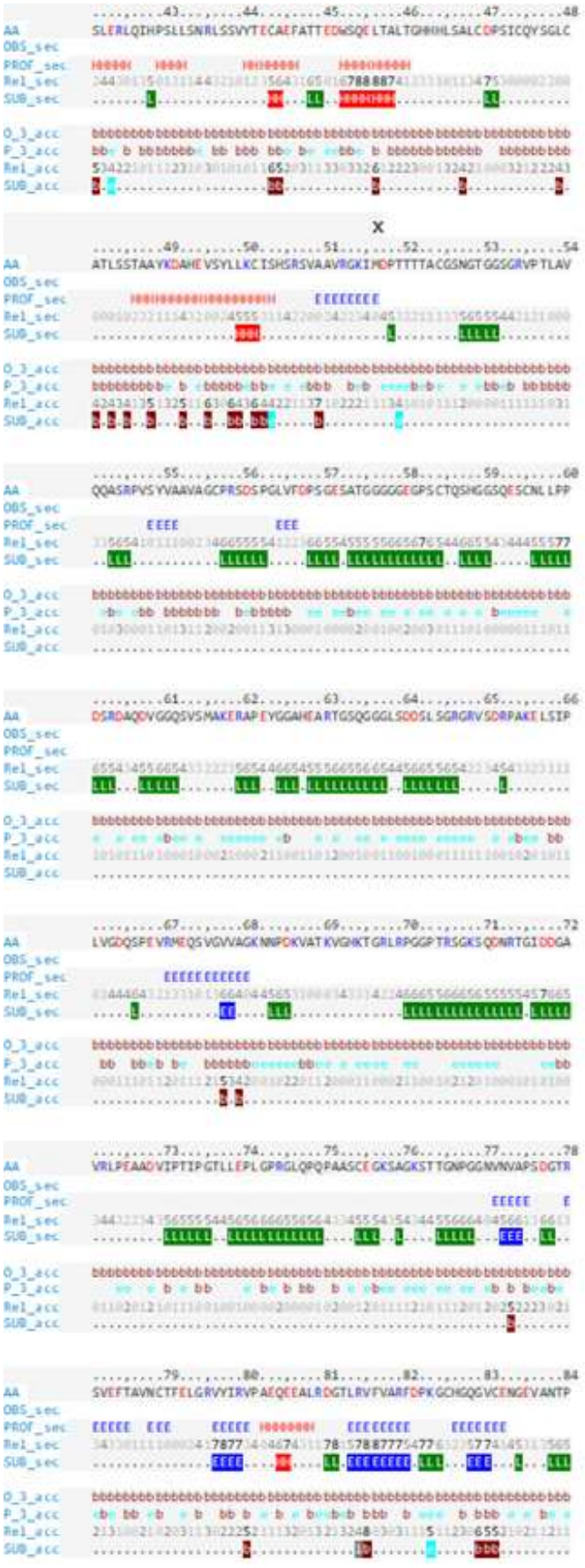
Kim SH, Ryabov EV, Kalinina NO, Rakitina DV, Gillespie T, MacFarlane S, Haupt S, Brown JWS, Taliansky M (2007). Cajal bodies and the nucleolus are required for systemic infection of a plant virus. EMBO Journal 26(8):2169-2179.

Kemenesi G, Földes F, Zana B, Kurucz, K., Estók, P, Boldogh S, Görföl T, Bányai K, Oldal M, Jakab F (2016). Genetic Characterization of Providence Virus Isolated from Bat Guano in Hungary. Genome

Announcements 4(3):e00403-16.
Luke GA, de Felipe P, Lukashev A, Kallioinen SE, Bruno EA, Ryan MD (2008). Occurrence, function and evolutionary origins of '2A-like' sequences in virus genomes. Journal General Virology 89(4):1036-1042.
Lupas AN, Gruber M (2005). The structure of alpha-helical coiled coils. Advances in Protein Chemistry 70:37-78.
McCartney AW, Greenwood JS, Fabian MR, White KA, Mullen RT (2005). Localization of the tomato bushy stunt virus replication protein p33 reveals a peroxisome-to-endoplasmic reticulum sorting pathway. Plant Cell 17:3513-3531.
Moore NF (1991). The *Nudarelia* β family of insect viruses (1991). Viruses of Invertebrates, 277-285. Edited by Kursak E, Mecel D, New York.
Panavas T, Hawkins CM, Panaviene Z, Nagy PD (2005). The role of the p33:p33/p92 interaction domain in RNA replication and intracellular localization of p33 and p92 proteins of cucumber necrosis tombusvirus. Virology 338(1):81-95.
Pringle FM, Johnson KN, Goodman CL, McIntosh AH, Ball LA (2003). Providence virus: a new member of the *Tetraviridae* that infects cultured insect cells. Virology 306(2):359-370.
Rajendran KS, Nagy PD (2004). Interaction between the replicase proteins of tomato bushy stunt virus *in vitro* and *in vivo*. Virology 326(2):250-261.
Rajendran KS, Nagy PD (2006). Kinetics and functional studies on interaction between the replicase proteins of tomato bushy stunt virus: requirement of p33: p92 interaction for replicase assembly. Virology 345(1):270-279.
Rajendran KS, Nagy PD (2003). Characterization of the RNA binding domains in the replicase proteins of tomato bushy stunt virus. Journal Virology 77(17):9244-9258.
Ryabov EV, Taliansk ME, Robinson DJ, Waterhouse PM, Murant AF, de Zoeten GA, Falk BW, Vetten HJ, Mark J Gibbs (2011). Umbraviridae. In: Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses, pp. 1191-1195. Edited by King AMQ, Adam MJ, Carstens EB, Lefkowitz EJ. San Diego: Elsevier Academic Press.
Scholthof K-BG, Scholthof HB, Jackson AO (1995). The tomato bushy stunt virus replicase proteins are coordinately expressed and membrane associated. Virology 208(1):365-369.
Short JR, Dorrington RA (2012). Membrane targeting of an alpha-like tetravirusreplicase is directed by a region within the RNA-dependent RNA polymerase domain. Journal General Virology 93:1076-1716.
Short JR, Knox C, Dorrington RA (2010). Subcellular localization and live-cell imaging of the *Helicoverpaarmigera*stunt virus replicase in mammalian and *Spodoptera frugiperda* cells. Journal General Virology 91:1514-1523.
Short JR, Nakayinga R, Hughes GE, Walter CT, Dorrington RA (2013). Providence virus (family: Carmotetraviridae) replicates vRNA in association with the Golgi apparatus and secretory vesicles. Journal General Virology 94:1073-1078.
Taliansky ME, Robinson DJ (2003). Molecular biology of umbraviruses; phantom warriors. Journal General Virology 84:1951-1960.
Taliansky M, Roberts IM, Kalinina N, Ryabov EV, Raj SK, Robinson DJ, Oparka KJ (2003). An umbraviral protein, involved in long distance RNA movement, binds viral RNA and forms unique, protective ribonucleo protein complexes. Journal Virology 77(5):3031-3040.
Taliansky ME, Robinson DJ, Murant AF (1996). Complete nucleotide sequence and organisation of the RNA genome of groundnut rosette umbravirus. Journal General Virology 77(9):2335-2345.
Tompa P, Csermely P (2004). The role of structural disorder in the function of RNA and protein chaperons. FASEB Journal 18(11):1169-1175.
Walter CT, Pringle FM, Nakayinga R, de Felipe P, Ryan MD, Ball LA, Dorrington RA (2010). Genome organization and translation products of Providence virus: insight into a unique tetravirus. Journal General Virology 91(11):2826-2835.
Walter CT (2008). Establishment of experimental systems for studying the replication biology of providence virus. Rhodes University. PhD thesis.
Wang Y, Zhang X, Zhang H, Lu Y, Huang H, Dong X, Chen J, Dong J, Yang X, Hang H, Jiang T (2012). Coiled-coil networking shapes cell molecular machinery. Molecular Biology of the Cell 23(19):3911-3922.

SUPPLEMENTARY FIGURE





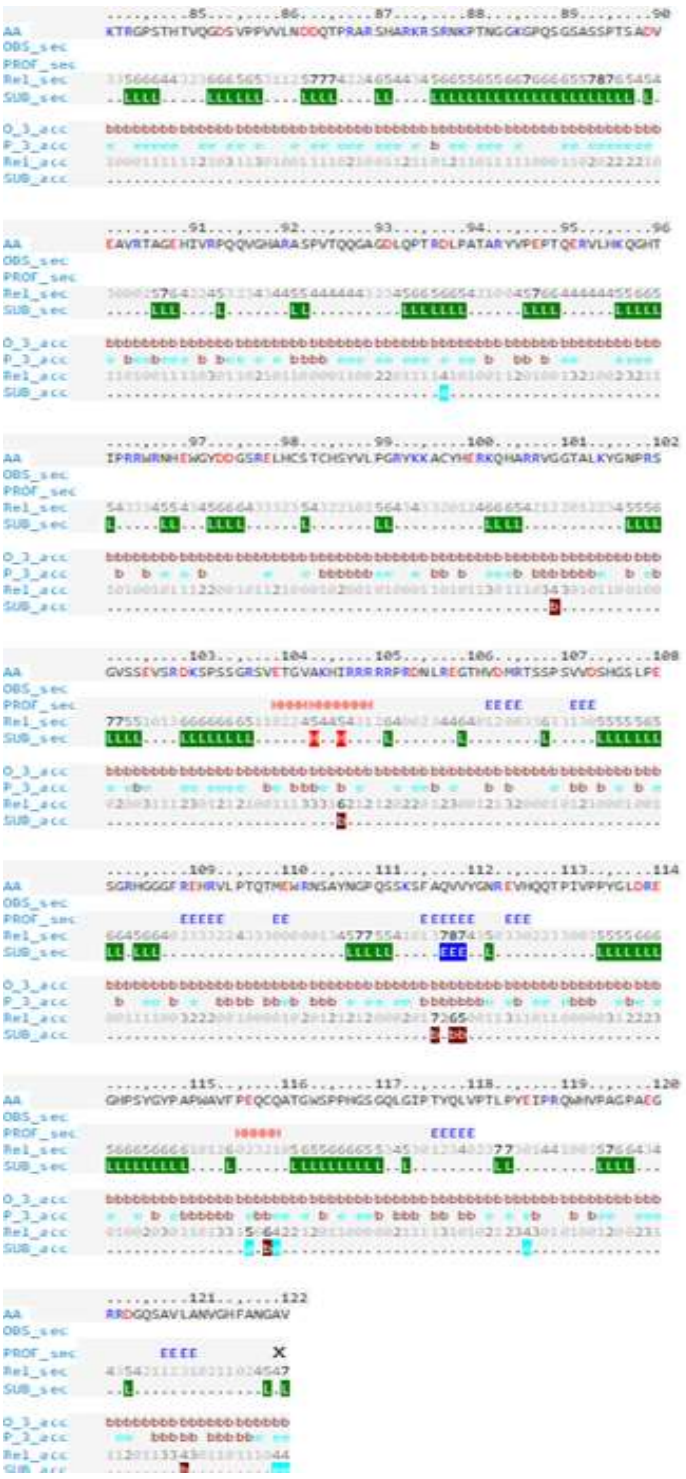


Figure S1. The predicted secondary structure in PrV p130. The complete amino acid sequence of PrV p130 showing the secondary structure type. The letter L shaded in green denotes loop structures while red and blue denote alpha helix and beta sheet structure types respectively. The crosses (x) above the amino acid sequence indicates the start and the end of the disordered region, spanning amino acids 517 to 1220. On the left is “AA’’ which represents amino acid sequence while “OBS_ sec, Prof_sec, Rel_sec, and Sub_sec represent methods incorporated in PredictProtein server for predicting protein secondary structures. The letters “e” and “b” denote solvent exposed residues.

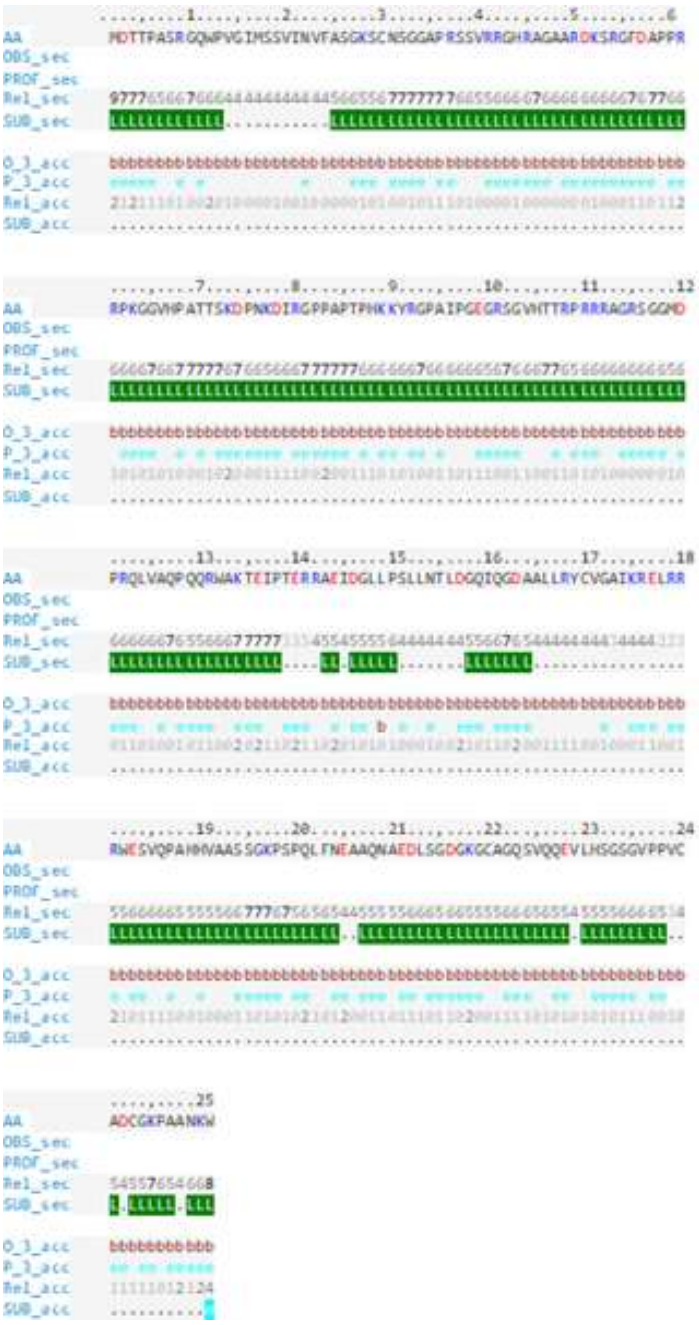


Figure S2. The predicted secondary structure of GrV p27. The complete amino acid sequence of GrV p27 showing the secondary structure types. The letter L shaded in green denotes loop structures. The letters “e” and “b” denote solvent exposed residues. The amino acid sequence of the protein is represented as “AA’’ while “OBS_ sec, Prof_sec, Rel_sec, and Sub_sec represent methods incorporated in PredictProtein server for predicting protein secondary structures.

```
seq        KPALVARAVALAMIPCREEVESLIDQQKNSASLWAVTNLLRSGWNSPWEW
frame-14   abcdefgabcdeabcaabcdefgabcdefgefgdefgfgefgabcdefgc
frame-21   bcdefabcdefgabcdefgabcdefgabcdefgdefgefgefgfgggfgc
frame-28   bcdefabcdefgabcdefgabcdefgabcdefgefgabcdefggefgabc
prob-14    ----------------3333333333333-------------------
prob-21    --------------------------------------------------
prob-28    --------------------------------------------------
```

Figure S3. The predicted sequence in PrV p40 that adopts the coiled-coil conformation. The partial sequence of PrV p40 showing amino acids, 317 – 330, with the potential of adopting a coiled-coil conformation.
The amino acid positions (a through g) of the heptad repeats are shown below the putative coiled-coil sequence that is detected by the window frame 14. The calculated probability is shown below the putative coiled-coil sequence at the same window frame.

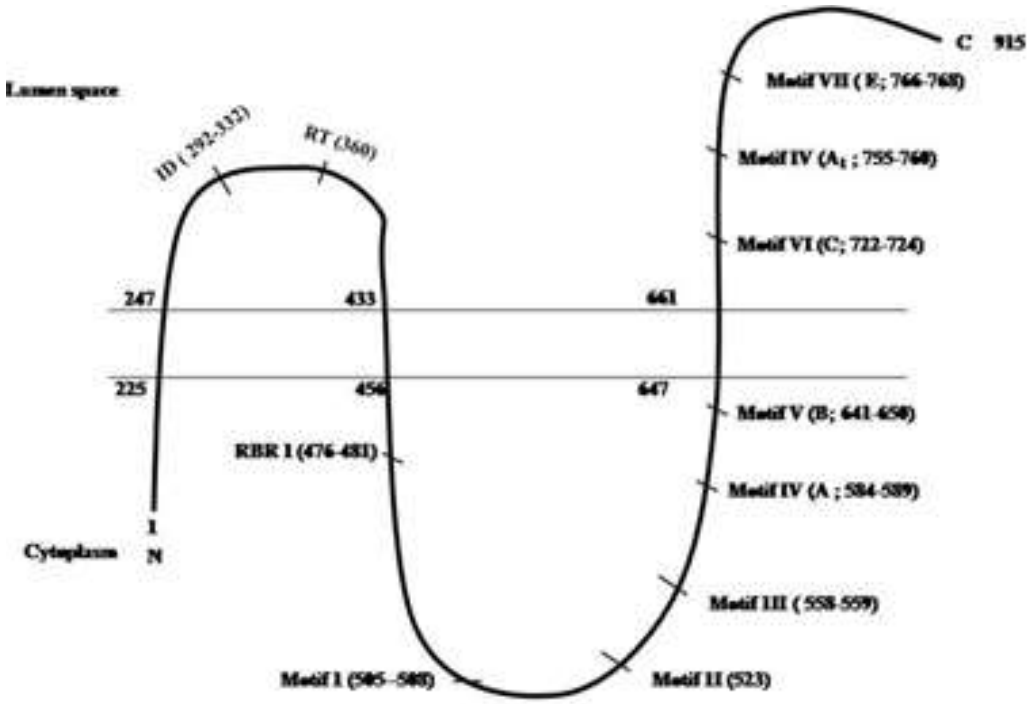


Figure S4. The hypothetical three transmembrane topology model of PrV p104 predicted by both TMpred and Kyte-Doolittle programs. The hypothetical lipid bilayer is depicted as an open space while the p104 amino acid sequence is depicted as a black line that traverses the hypothetical lipid bilayer three times at positions 225 - 247, 433 - 456 and 647 – 661 respectively. The amino acid stretch of each transmembrane helice is annotated at either side of the lipid bilayer. The RdRp motifs I, II, III, IV (A), V (B), VI (C), IV (A1), VII (E) are labeled on the protein, each with its amino acid position relative to the p104 amino acid sequence. The putative interaction sequence is denoted as ID and RNA binding region is denoted as RBR 1. The readthrough stop codon is denoted as RT.

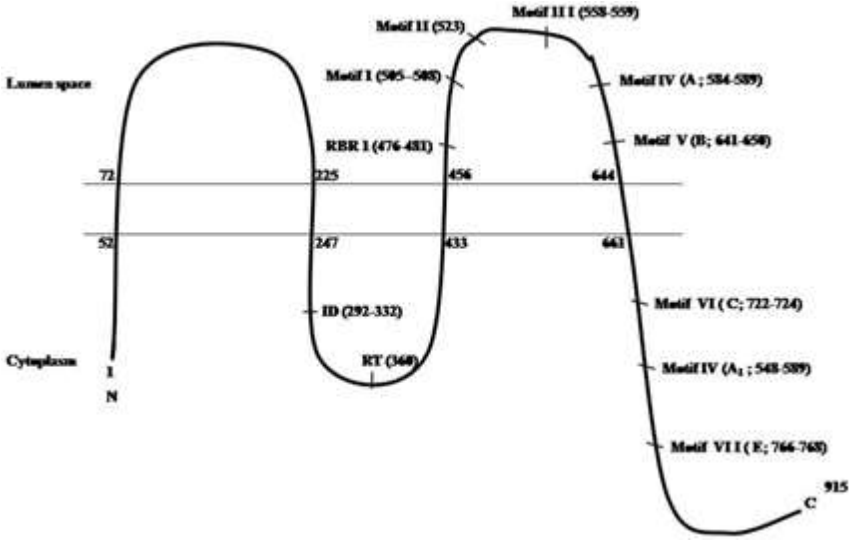


Figure S5. The hypothetical four transmembrane topology model of PrV p104 predicted by TMpred and Kyte-Doolittle programs. The hypothetical lipid bilayer is depicted as an open space. The p104 amino acid sequence is depicted as a black line traversing the hypothetical lipid bilayer four times at positions 52 - 72 (predicted by TMpred), 225 - 247, 433 - 456 and 647 - 661 (predicted by TMpred and Kyte-Doolittle) programs. The amino acid stretch of each transmembrane helice is annotated at either side of the lipid bilayer. The RdRp motifs I, II, III, IV (A), V (B), VI (C), IV (A1), VII (E) are labeled on the protein, each with its amino acid position relative to the p104 amino acid sequence. The putative interaction sequence is denoted as ID and RNA binding regions as RBR 1. The readthrough stop codon is denoted as RT.