

SCIENTIFIC REPORTS



OPEN

Prion-like Domains in Eukaryotic Viruses

George Tetz  & Victor Tetz

Prions are proteins that can self-propagate, leading to the misfolding of proteins. In addition to the previously demonstrated pathogenic roles of prions during the development of different mammalian diseases, including neurodegenerative diseases, they have recently been shown to represent an important functional component in many prokaryotic and eukaryotic organisms and bacteriophages, confirming the previously unexplored important regulatory and functional roles. However, an in-depth analysis of these domains in eukaryotic viruses has not been performed. Here, we examined the presence of prion-like proteins in eukaryotic viruses that play a primary role in different ecosystems and that are associated with emerging diseases in humans. We identified relevant functional associations in different viral processes and regularities in their presence at different taxonomic levels. Using the prion-like amino-acid composition computational algorithm, we detected 2679 unique putative prion-like domains within 2,742,160 publicly available viral protein sequences. Our findings indicate that viral prion-like proteins can be found in different viruses of insects, plants, mammals, and humans. The analysis performed here demonstrated common patterns in the distribution of prion-like domains across viral orders and families, and revealed probable functional associations with different steps of viral replication and interaction with host cells. These data allow the identification of the viral prion-like proteins as potential novel regulators of viral infections.

Recently, prions and their infectious forms have attracted a lot of research attention^{1,2}. The infectious prion forms (PrP^{Sc}) represent the misfolded normal proteins (PrP^C) and were shown to be infectious, since they can self-propagate and interact with the endogenous PrP^C, catalyzing their conversion into pathological PrP^{Sc}^{3–7}. Previously they had been primarily known as the inducers of transmissible spongiform encephalopathies, however, today they have been shown to be involved in the development of a variety of neurodegenerative diseases^{8–10}.

Recently, the abnormal conformation of self-propagating PrP^{Sc}s was found to be associated with the formation of the toxic, misfolded, insoluble, and highly-ordered fibrillar cross- β aggregates of β -amyloid, *tau*, and TDP-43 proteins in Alzheimer's disease and amyotrophic lateral sclerosis^{11–14}. The follow-up studies demonstrated that the pathological protein conversion and the deposition of the insoluble protein aggregates are associated with the development of other diseases, including Parkinson's, Huntington, fatal familial insomnia, ataxias, diabetes, and others^{9,15,16}.

However, protein misfolding was shown to play important physiological roles as well in eukaryotes and prokaryotes^{17–21}. Self-perpetuating properties of prions are important for the formation of bacterial and fungal biofilms, bacterial bacteriocin functioning, molecular transport and secretion, and the preservation of long-term memory in yeasts^{22–25}. Moreover, prions were recently shown to participate in the communication between prokaryotes and eukaryotes, resulting in the alterations in *Caenorhabditis elegans* amyloid formation following its colonization with amyloid-producing *Escherichia coli*²⁶.

Although the molecular mechanisms underlying *de novo* prion formation remains elusive, the aggregation of PrPs is an amino-acid sequence-dependent process. Most prions contain specific domains enriched in asparagine (Q) and glutamine (N), which, together with the average residue hydrophobicity and net sequence charge, allowed the development of algorithms for the identification of candidate prionogenic domains (PrDs) based on the hidden Markov model (HMM)^{20,27–30}.

The HMM is currently used in many bioinformatic approaches for the statistical representation of prion domains, which allow, using the probabilistic sequence model of maximum likelihood estimation, to evaluate the compositional similarity of proteins and prions. One of these approaches is prion-like amino acid composition (PLAAC) analysis, which allows the evaluation of proteins containing PrDs, defined as domains with the compositional similarity to yeast prion domains, based on amino-acid interactions^{27,31}. The resulting log-likelihood

Human Microbiology Institute, New York, NY, 10027, USA. Correspondence and requests for materials should be addressed to G.T. (email: georgetets@gmail.com)

ratio (LLR) indicates the possibility that the analyzed protein is a prion. Using PLAAC algorithms, PrDs defined as domains shown to contain at least a domain compositionally similar to yeast prions, have recently been investigated in different eukaryotic and prokaryotic species, confirming their important regulatory and functional roles^{20,32–34}. There are other algorithms, such as PAPA and PrionW, using an experimentally derived prion propensity score combined with explicit consideration of the intrinsic disorder, that help to predict prion domains bioinformatically^{35–38}. Recently, we investigated the PrDs in phagobiota and determined that these domains can be found in bacterial and archaeal virus families, which increased our understanding of their possible interplay with microbiota and implication for human health³⁹.

Similar to the bacterial viruses, eukaryotic viruses are found in nearly all ecosystems and they infect different types of organisms, including animals, insects, protists, and plants, but their life cycle is similar in different organisms, comprising the attachment, entry, biosynthesis of viral nucleic acids and proteins, maturation, and release of progeny⁴⁰. The nature of the viral replication cycle leads to their pathogenicity, although certain viral species induce persistent infections⁴¹. Viruses are efficiently disseminated by horizontal and vertical transmission, and they are the causative agents of many devastating diseases, such as flu and some cancers^{42,43}. However, the detailed molecular mechanisms underlying the pathological processes have not been completely elucidated yet^{14,44}.

Moreover, despite previous efforts, the presence of PrDs in eukaryotic viruses has not been described very well and the PrDs have been identified in only several viral families^{19,34}. Therefore, the PrDs distribution in different viral families and species and their functionalities have not been determined to date.

Here, we performed a detailed study of the putative prion domains in all known eukaryotic viruses. Using an HMM algorithm, we retrieved all available eukaryotic viral protein sequences from the UniProt KB database⁴⁴. To the best of our knowledge, this is the most extensive effort aimed at the identification of candidate PrD sequences among eukaryotic viruses. Furthermore, we analyzed the regularities in the distribution of PrDs in different viral taxa, correlation of this distribution with viral structure, viral hosts, and protein functions. The PrDs were identified using different algorithms, including Gene Ontology (GO)⁴⁵. Our results may contribute to the better understanding of the host-viral interactions and the relationship between viral prions and pathogenicity.

Materials and Methods

Protein sequences. To identify the PrDs present in viral proteomes, protein sequences were obtained from the UniProt KnowledgeBase (Swiss-Prot and TrEMBL). Protein functions were predicted using the GO terms and manually curated using the information from the UniProt database, the National Center for Biotechnology Information (NCBI), and the literature data⁴⁶.

Identification of PrDs in viral proteomes. The presence of PrDs in viral proteomes was analyzed in the known viruses, excluding bacteriophages, using the PLAAC prion prediction algorithm, based on the HMM, and the identification of PrDs was based on the compositional bias towards asparagine and glutamine aminoacyls, an average residue hydrophobicity, and the net charge of sequences. The output probabilities for the PrDs states in the PLAAC were constructed based on the amino-acid frequencies in the PrDs of *Saccharomyces cerevisiae*. Consequently, this basis can be altered, using the parameter “Alpha value”, which allows a continuous interpolation between organism-specific background frequencies (Alpha = 0.0) and *S. cerevisiae* background frequencies (Alpha = 1.0). Here, we used Alpha = 0.0, representing species-independent scanning, to identify the PrDs.

For the analysis, we have adjusted the total number of viral proteins contained in the UniProt database, since in the proteomes of different viruses, multiple fragments of the same proteins had multiple representation. Therefore, multiple copies of the same sequences were removed in Excel (Windows 10) using the ‘remove duplicates’ function. We used a low LLR cutoff of 0.003, in order to analyze the majority of PrDs and their distribution among different viral orders and families, and 2,681 PrDs have been identified (Supplementary Table 1). Prion-like domains of top 100 scoring PrDs in different viral species were also predicted by the program PAPA using default values and a defined cutoff score of 0.05 for prediction of the prion versus non-prion proteins (Supplementary Table 6)^{35,36}. The regularities in the likelihood of the identified PrDs to be prions, and their distribution among different viral orders and families were analyzed. The functions of proteins with the identified PrDs were classified using the manually-curated GO categories and were based on the major steps of viral replication. A heatmap was generated using R-statistical computing (www.r-project.org) with the “levelplot” package. The values in the heatmap range between the lowest (blue) and the highest (red) LLR values.

Statistical analysis. All statistical analyses were conducted using package Statistica for Windows (version 5.0) (StatSoft, Inc.). Data were compared between the viral orders, families, and species by using a χ^2 test or the Fisher’s exact test. To detect differences in multiple comparisons, one-way analysis of variance (ANOVA) was fitted with the standard confidence interval of 95%. All results were considered statistically significant for $p < 0.05$.

Results

Identification of PrDs in viral proteomes. Using the prion-prediction PLAAC algorithm, we identified 2,679 PrDs in proteins from 735 different viruses. In total, we analyzed 2,742,160 proteins derived from the UniProtKB database from over 3000 known viral species (the whole virome discovered up-to-date) according to the International Committee on Taxonomy of Viruses (ICTV) (Supplementary Tables 1, 2)⁴⁷.

The average numbers of LLRs varied between the DNA and RNA containing viruses as well as between the enveloped and non-enveloped ones. PrDs were more frequently found in the DNA-containing viruses. Enveloped viruses were also more frequently found to harbor PrD compared with the non-enveloped ones (Fig. 1).

High levels of PrDs were found in *Herpesvirales*, *Megavirales*, *Mononegavirales*, *Nidovirales*, *Picornavirales*, and *Tymovirales* (Fig. 2) (the members of the unassigned viral orders, represented by different unrelated families,

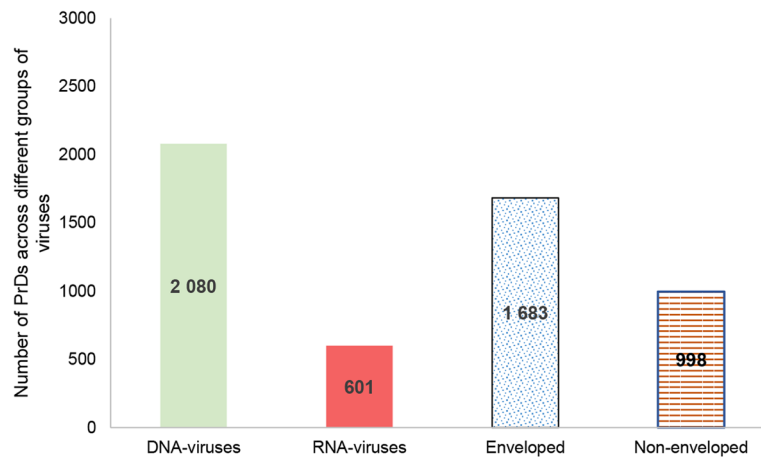


Figure 1. PrD enrichment in the proteome of different viruses. Numbers insides bars represent the total numbers of PrDs identified in each group.

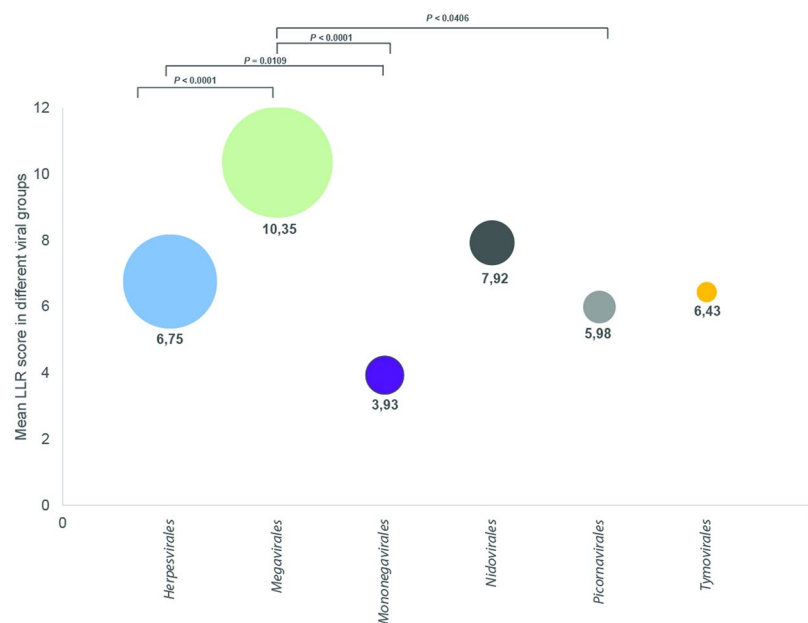


Figure 2. PrD enrichment in viral proteomes and the LLR scores. The ratio between PrD-containing proteins in each group and the total number of viral proteins is presented. Numerical values are medial LLR scores of the PrDs, and the circle size indicates the number of identified PrDs. Data were analyzed using one-way ANOVA.

Order	Number of PrD-containing species within one order	Total number of species within an order	PrD-containing species as the percentage of the total species number	P-value
<i>Herpesvirales</i>	74	103	71,84%	<0.0001
<i>Megavirales</i>	78	ND	ND	ND
<i>Mononegavirales</i>	35	212	16,51%	<0.0001
<i>Nidovirales</i>	60	64	93,75%	<0.0001
<i>Picornavirales</i>	44	138	31,88%	0.7579
<i>Tymovirales</i>	16	179	8,94%	0.948
Unassigned	427	2467	17,31%	<0.0001

Table 1. PrD enrichment in the proteomes of different viral orders.

Order	Number of PrD-carrying viral species	Total number of PrDs identified in the order	Mean number of PrDs per species
<i>Herpesvirales</i>	74	500	6.75
<i>Megavirales</i>	78	694	8.86
<i>Mononegavirales</i>	35	85	2.42
<i>Nidovirales</i>	60	114	1.90
<i>Picornavirales</i>	44	60	1.36
<i>Tymovirales</i>	16	23	1.43
Unassigned	427	1204	2.83

Table 2. Mean PrD numbers per species in the same viral order.

are presented in Suppl. Table 2). The distribution of PrDs was shown to vary, with the highest prevalence found in *Herpesvirales* (LLC = 6.54).

To analyze the presence of PrDs in different viral orders, we evaluated the ratio between the species identified in this study to possess at least one PrD and the total number of different viral species within that order (Table 1)⁴⁷.

We found that the highest number of PrD-containing species are found among *Nidovirales* and *Herpesvirales*, with over 93.75% and 71.84% of species, respectively, containing PrDs, while the lowest numbers were found in *Tymovirales*, with only 8.94% of species with identified PrDs. We have not included the results of *Megavirales* analysis due to the lack of classification data for this novel viral order⁴⁸.

Furthermore, we calculated the mean number of PrDs per species, as the ratio of the total number of PrDs identified in viral proteomes attributed to an order to the total number of PrD-bearing species identified in this order. The highest average numbers of PrDs per species were identified in *Megavirales* and *Herpesvirales* species (Table 2; Supplementary Table 3).

Next, we evaluated the LLRs in the viral orders and families. The largest number of viruses with the highest LLR scores, over 50 and 40, were identified in the order *Megavirales* (families *Mimiviridae*, *Phycodnaviridae*, and *Poxviridae*), while only a few were obtained in *Herpesviridae*. (Supplementary Tables 4 and 5). By analyzing top 100 scoring PrDs of the viruses with the greatest prion-forming potential, we evaluated the highest LLR scores predominantly among *Megavirales*, *Herpesviridae*, and in viruses of unassigned orders (Supplementary Table 6). Twenty seven percent of these top 100 PrDs were identified in the *Mimiviridae* species, order *Megavirales*, of *Acanthamoeba*, with the mean LLR score of 48.68. We also applied the PAPA prion prediction algorithm to these top 100 scoring PrDs and the majority of the results were consistent with the PLAAC analysis^{35,36}.

Additionally, we analyzed the PrD enrichment in the proteomes of different viral species. The highest enrichment rate was found for the members of the *Megavirales* order, with at least five PrDs per proteome in the viruses belonging to the *Mimiviridae* and *Phycodnaviridae* families (Supplementary Table 7). The highest number of different viral species with over 10 PrDs per proteome was found in the *Herpesviridae* family.

Association of viral PrDs with the functional domains. We clustered PrDs into six functional groups based on the major processes during the viral interaction with the host cell: adsorption and entry, biosynthesis, including the transcription, translation, and synthesis of viral components, maturation, assembly, release, and a group comprising proteins with an unknown function⁴⁹. We separately analyzed the PrDs in the viral precursor proteins⁵⁰. Additionally, we analyzed the PrDs identified in proteins with the functions not related to the main viral processes, but that, nevertheless, play important roles in disease pathogenesis, the virus-induced changes in the morphological, biochemical, or growth parameters of cells, and the suppression of host complement activation. The correlations the PrDs and protein functions were identified, and the PrD numbers, their LLR scores, and viral families were analyzed (Fig. 3).

To facilitate the interpretation of the results, we grouped the proteins based on their functions using the GO terms (Supplementary Figs 1–10).

Following this, we identified 433 PrDs (medium LLR score, 5.05) in proteins involved in the viral adsorption and entry, and predominantly associated with the host cell-membrane binding. This group contains proteins belonging to different GO terms, including the integral component of membrane, viral envelope, virion attachment to host cell, fusion of virus membrane with host plasma membrane, receptor-mediated virion attachment to host cell, and others (Supplementary Fig. 1). We identified PrDs in proteins associated with the adsorption and those involved in the direct contact with the host cell, such as spike proteins, VP1, glycoproteins, hemagglutinin-neuraminidase, etc.⁵¹. Heatmap analysis results showed that the members of *Baculoviridae* and *Herpesviridae* have the highest number of PrDs associated with the viral adsorption and entry (Fig. 3). Furthermore, we identified PrDs in glycoproteins and membrane proteins of viruses that affect human health, such as human α -, β -, and γ -herpesviruses (human herpes virus 1, 2, 5, and 7) and other viruses associated with human diseases, such as hepatitis B and C, Marburg virus, rotavirus A, human immunodeficiency virus 1 (HIV 1), and others⁵².

The biggest cluster of PrDs (502 proteins) contained the proteins involved in viral transcription, translation, and protein synthesis (LLR score, 6.69), with multiple molecular functions and belonging to different GO terms. The members of *Herpesviridae* family contained the majority of these PrDs (Fig. 3; Supplementary Table 1; Supplementary Fig. 2). We identified PrDs in the DNA polymerases of different human herpesviruses, such as cytomegalovirus, Epstein-Barr, varicella-zoster viruses, and herpes simplex virus 2. Additionally, we detected

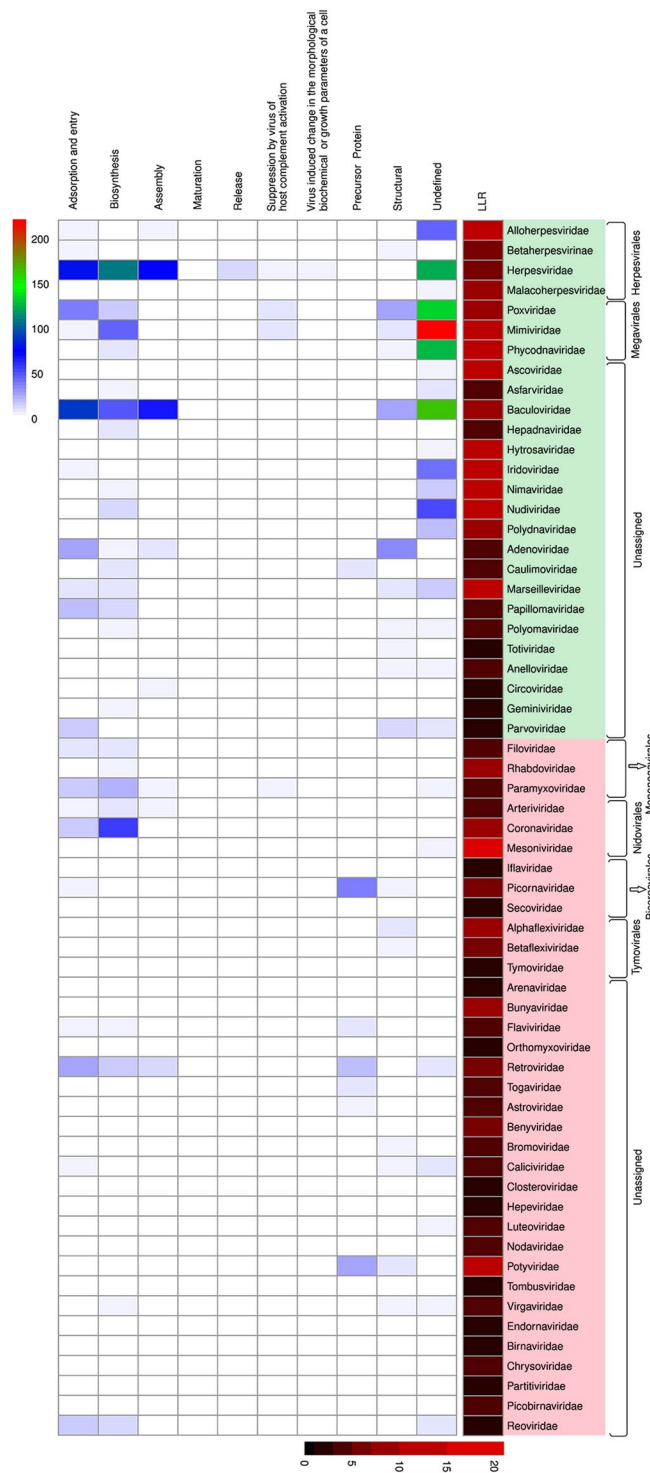


Figure 3. PrD distribution in viral proteins and the GO term analyses. Heatmap of PrD distribution in viral proteins. The correlations between the functions of PrD-containing proteins, PrD numbers, their LLR scores, and viral families are presented. Columns, the main protein function; rows, viral families. Cells indexed by rows and columns are marked using a color gradient, ranging from white (no PrD-containing proteins) to red (maximum number of proteins with PrDs). Mean LLC scores of proteins presented in the far-right column are denoted by using a gradient of red, ranging from black (score 0) to saturated red (score 70; color bar). Viral families are grouped according to their genetic material. Green background, DNA-viruses; red background, RNA-viruses.

them in the Epstein-Barr nuclear antigens (EBNA) and large tegument protein deneddylase of these viruses, in the RNA-directed 5'-3' RNA polymerases and nucleoproteins of *Filoviridae* viruses, such as Marburg virus and Zaire ebolavirus, in the nucleoproteins of human coronavirus and porcine epidemic diarrhea (PED) virus (*Coronaviridae*), and others^{53,54}.

Following this, we analyzed PrD-containing viruses associated with the viral assembly. We identified 209 PrDs with the mean LLR score of 7.79. The main GO terms represented were the viral capsid assembly, serine-type endopeptidase activity, nuclear capsid assembly, viral DNA genome packaging, and others (Supplementary Fig. 3). The key PrD-containing proteins shown to be involved in the viral assembly were identified in the *Baculoviridae* and *Herpesviridae* families. We identified desmoplakin as the main PrD-containing protein in *Baculoviridae*, capsid scaffold protein and small capsomere-interacting protein 1 were the most abundant in different herpesviruses, Gag protein in many *Retroviridae* and other viruses (Fig. 3)^{55,56}.

The identified PrDs in proteins involved in the release of viral progeny from the host cell were shown to be less abundant, with only 19 proteins found to contain these domains (LLR score, 3.68). In the GO terms, this group predominantly comprised proteins associated with the DNA packaging and viral release from the host cell. The highest number of them were identified in *Herpesviridae*, including partial proteins, capsid vertex component 2, and tegument protein pp150 (Fig. 3; Supplementary Fig. 4).

Additionally, we identified six PrDs in proteins associated with the viral maturation (LLR score, 23.61) and with the GO terms associated with the integral components of the membrane and methyltransferase activity in different viruses (Fig. 3; Supplementary Fig. 5)⁵⁷.

The PrDs were also detected in 223 structural proteins, predominantly represented by capsid, coat proteins, and hexons (mean LLR score, 5.78)⁵⁸. Notably, the majority of these proteins were found in different non-enveloped viruses, primarily from the *Adenoviridae* and *Baculoviridae* families, and these were less abundant in the enveloped viruses, primarily belonging to *Poxviridae* (Supplementary Fig. 6).

Furthermore, we identified 138 PrDs (mean LLR score, 6.47) in the viral precursor proteins. Positive-strand RNA viruses are characterized by a positive strand RNA genome encoding a single poly-protein precursor, which, during the post-translational processing, are cleaved and processed into the mature proteins. We identified PrDs in the genome polyproteins of *Picornavirales* (foot-and-mouth disease virus, enterovirus B, and cardiovirus B) and *Flaviviridae* (Zika virus, hepatitis C virus), in the Gag polyprotein of *Retroviridae* (HIV1, bovine leukemia virus), and others (Fig. 3, Supplementary Table 7)^{59,60}.

The PrDs identified in the proteins associated with the viral suppression of host complement activation were less abundant, and only 39 of these proteins were identified (mean LLR score, 7.11). In the GO terms, they were represented with the G-protein coupled receptor activity, evasion or tolerance of host immune response, metal ion binding, and unassigned processes. PrDs were found in NF-kappa B inhibitors, envelope glycoprotein UL33, ankyrin repeat-containing protein, and others, and among different viruses, including some important human pathogens, such as cytomegalovirus, Kaposi's sarcoma-associated herpesvirus, and HIV1 and 2 (Supplementary Fig. 8)^{61,62}.

We found 14 PrDs (mean LLR score, 11.09) in proteins implicated in the virus-induced change in the morphological, biochemical, or growth parameters of cells. Among these, we identified late membrane protein 1 and K1 in Epstein-Barr virus and Kaposi's sarcoma-associated herpesvirus (Fig. 3; Supplementary Fig. 9)⁶³.

Finally, we identified 1097 PrDs in proteins with still unknown functions (mean LLR score, 9.79). The vast majority of these are uncharacterized proteins, which has still not been reviewed in the GO terms (Supplementary Fig. 10).

Discussion

This study is the most complete evaluation of PrDs among viruses except for the bacteriophages³⁴. The results of our study highlight some previously overlooked viral characteristics that may play important roles in viral infections. We determined that more PrDs can be found in the DNA-viruses compared with their numbers in the RNA-viruses, and in the enveloped viruses, compared with that in the non-enveloped ones. This may be partially explained by the larger genome size and protein numbers in DNA-viruses⁶⁴.

We identified PrDs in functionally distinct proteins of different viral orders, indicating that these PrDs are conserved in different viruses. However, the PrDs were not identified in all viral families and species. Our analyses demonstrated that only approximately 23% of all analyzed viral proteomes available in public databases contain at least one PrD, suggesting that the presence of PrD-containing proteins may be beneficial, but not obligatory. We identified PrDs in many human viral pathogens, but other viruses affecting human health were shown to have a few or no PrDs in their proteomes, such as hepatitis A, E, and D viruses, papillomaviruses, some members of *Orthomyxoviridae*, and others.

At the order level, we showed that the PrDs are more frequent among *Megavirales* and *Herpesvirales*, while, at the species level, the highest number of PrDs was found in *Acanthamoeba polyphaga* mimivirus, *Paramecium bursaria* Chlorella virus NY2A, *Acanthamoeba castellanii* mamavirus (*Megavirales*), and *Heliothis zea nudivir* (unassigned order). Among human pathogens, the highest prevalence of PrD was found in cytomegalovirus and Epstein-Barr virus (*Herpesvirales*) and HIV1 (*Retroviridae* family, unassigned order) (Supplementary Table 7).

The analysis of the top 100 scoring PrDs, with the highest number of QN-rich domains, they were found to be most common among *Mimiviridae*, which infect *Acanthamoeba*, and *Phycodnaviridae*, which infect algae and belong to the *Megavirales*. Of these, only some proteins were *Herpesvirales* proteins, while the majority of them was shown to be identified in the viruses of the unassigned order. No human viruses were shown to have LLR scores over 31 and none were represented in the top 100 LLR-scoring group. The majority of these proteins has not been characterized, with still unknown functions, and therefore, the functional relevance of these findings remains unclear. We also applied the PAPA prion prediction algorithm to these top 100 scoring PrDs^{35,36}. The

majority of data received by the PAPA, was overlapping with PLAAC algorithm, which is in agreement with previous studies, showing that PLAAC and PAPA commonly overlap as both programs were trained on prion proteins of *S. cerevisiae*³⁸. However, there were certain differences, and some PrDs that had high PLAAC score were found to be below the 0.05 cutoff of PAPA algorithm. Such a discrepancy can be explained by the fact that although both of these programs have a high level of prediction accuracy for prion and non-prion protein determination, the precision of the analysis is below 100%^{35,36}.

The order *Megavirales* is a recently established order that comprises of diverse group of the DNA-viruses infecting eukaryotic hosts, which are characterized by large genomes, almost 10 times larger than those of the *Herpesviridae*⁴⁸. Here, DNA-viruses were found to harbor more high-scoring prions, as expected, but the high LLR scores obtained for these viruses is not due to the longer amino-acid sequences, but to the increased presence of QN-residues.

Furthermore, we aimed to determine the correlation between the PrD-containing protein functions and the frequency of PrDs in the viral proteomes found in different viral families. Adhesion and entry of viral nucleic acids represent crucial steps in the viral-host interactions and the viral PrD-containing proteins showed to be involved in these processes represented the second largest group. We identified the PrDs in the viral surface proteins that are involved in the direct contact and fusion of viruses with the host cell membrane, indicating that PrDs may be functionally implicated in these processes as well. Moreover, a similar trend was previously noted in the distribution of PrDs in proteins responsible for the bacterial and bacteriophage interactions³⁹. Interestingly, most PrDs were found in *Baculoviridae*, the rod-shaped viruses. We identified PrDs in 56 out of 66 known species belonging this family, indicating that this high prevalence of PrDs in one viral family may not be a coincidence. Therefore, we further showed that one of the PrDs associated with the cell adhesion and entry most frequently found in *Baculoviridae* is occlusion-derived virus envelope protein 66 (ODV-E66), which was recently the first identified viral chondroitin lyase, an enzyme that degrades chondroitin and is associated with viral entry^{65,66}.

Of 543 PrDs found to be associated with the viral interaction with the host cells, only four proteins were identified in the plant viruses (potato mop-top virus, Dasheen mosaic virus, only Syngen Nebraska virus 5, and Fiji disease virus), while none were identified in the fungi viruses. This may be explained by the presence of cell wall in the plant and fungal cells, requiring different mechanism of viral entry⁶⁷. Plant viruses are known to have no specific mechanisms of entry, but instead they take advantage of the plant injury, vectors such as insects, or through a cell-to-cell movement of viral progeny in the infected plant^{68,69}. Taken together, the presented findings, showing that PrDs are present in the proteins of animal viruses that interact with cell membranes, and their absence from the plant and fungal viruses indicates that the identified PrDs associated with adhesion and entry may have important functional roles.

Proteins involved in the viral biosynthesis were shown to harbor PrDs as well. We found numerous PrDs in different nucleic acid-binding proteins, proteins with kinase activity, regulators of chromosome condensation, and others. Several PrDs were found in the DNA and RNA polymerases, helicases, and EBNA1. However, the roles of PrDs in the viral proteins associated with biosynthesis remain unclear, but they may be necessary for the efficient protein-nucleic acid interactions. Similar trend was observed in the eukaryotes, where the most abundant PrDs can be found in the nucleic acid-binding proteins⁷⁰. In humans, the QN-biased amino acid-enriched regions are frequently found in the RNA-binding proteins and regulatory molecules⁶⁶. These nucleic acid-binding prion candidates are found in the proteins associated with neurodegenerative disorders, such as amyotrophic lateral sclerosis (ALS), Alzheimer's disease, and Huntington's disease, and they have been suggested to act as the epicenters of misfolding⁷¹.

Some viruses may be implicated in the prion misfolding in humans since it was observed that the *de novo* appearance of prions can be facilitated by another PrD-containing protein⁷². Previously, human herpes virus was shown to be associated with the development of Alzheimer's disease, and several epidemiological studies demonstrated the presence of HSV1 antigens in the cerebrospinal fluid of Alzheimer's disease patients^{73,74}. It can be speculated that the presence of multiple PrDs identified in this study in the HSV1 proteins, including nucleic acid-binding and surface proteins, may represent a seeding trigger of the protein misfolding in Alzheimer's disease. However, further experiments are required to confirm this.

Moreover, we detected some interesting trends in the PrD distribution among other functional groups of viral proteins. The presence of PrDs in viral assembly proteins is most likely important for the capsid assembly, which is known to be characterized by the nucleation-and-growth mechanism, dependent on the capsid assembly kinetics⁷⁵. Therefore, the glutamine-rich motifs, which mediate protein-protein interactions and have greater potential for the kinetic conformational diversity, in capsid assembly proteins may determine the pathways of capsid protein complex assembly⁷⁶. The majority of PrDs in the assembly proteins were found in the enveloped viruses, while PrDs in the structural proteins were predominantly identified in the non-enveloped ones. Further analyses indicated the PrD enrichment in the assembly proteins, such as adenoviral major capsid proteins hexons, found in different members of *Adenoviridae*, including human adenovirus serotype 40 (AdV40)⁷⁷.

Adenoviral hexons are known to play multiple roles including structural roles and those associated with the host-cell interactions. The identified PrDs, according to PLAAC analysis, are located within 130–167 amino acids in the AdV40, which partially corresponds to a variable region V1 located in the AdV40 serotype-specific loop 1^{78,79}. These hypervariable regions are located on the surface of hexons, and the loops protruding from the capsid are thought to be implicated in the immune response triggering through the interactions with the human adenovirus-specific T-cells^{68,80,81}.

Another thing that caught our attention in terms of viral assembly and prionogenic properties is that in eukaryotes, PrDs are known to be implicated in liquid-liquid phase separation (LLPS)^{82–85}. LLPS is known to be an important process in the nucleation and growth of protein crystals and is suggested as the first step for viral capsid growth^{86,87}. Therefore, the discovery of PrDs in proteins associated with viral capsid growth indicates a possible role of these prionogenic properties in viral assembly. An indirect reference to that is that human cytomegalovirus capsid scaffolding

protein, which plays an essential structural role in assembling the viral capsid, was identified in the current study as having a PrD, and in recent work by Vernon, R. *et al.*, it was shown to be associated with phase separation⁸⁸.

In viruses, the presence of PrDs can be associated with the precursor proteins of positive-sense RNA viruses, such as *Picornavirales* and *Retroviridae*. By analyzing this group, we detected the PrDs in some precursor proteins, which were not found in the functional proteins cleaved from these primary polyproteins of *Picornavirales*. Only enterovirus B were shown to have the PrDs in both polyprotein and capsid protein VP1. These observations require additional investigations, since they cannot be due to an incomplete analysis, as the UniProt database contains the data on many *Picornavirales* proteins. It is likely that certain post-translational modifications of precursor polyproteins may result in the removal of the PrDs during the process of maturation.

Furthermore, PrD-containing proteins were found to be associated with the suppression of host complement activation and virus-induced changes in cells, including the modulation of host apoptotic process⁸². We identified the PrDs in these proteins in many unrelated viruses of different hosts. The PrDs were primarily identified in human and insect viruses with demonstrated ability to establish persistent infections: *Baculoviridae*, HIV1, and *Herpesviridae* species, such as HSV1, Epstein-Barr virus, cytomegalovirus, and oncogenic HSV^{89,90}. This indicates that the PrDs in these proteins may be implicated in the establishment of persistent viral infections and affect the adaptive immune response.

Taken together, we identified numerous putative PrD-containing proteins in viruses. Although we used a low threshold of PLAAC score for identification (or in other words even proteins with a low probability of being a prion were included in the analysis), there were still over 650 proteins with high LLR score over 10, which were the most promising prion candidates (Supplementary Table 8). We observed consistent PrD distribution patterns in different viral families and species, and these domains were identified in a variety of proteins. However, since the majority of viruses were shown to lack the PrDs, this shows that the presence of PrDs is beneficial, but not obligatory, which agrees with the results obtained for the PrDs found in bacteria and bacteriophages^{32,34}. Further analyses are required to elucidate the role of the identified PrDs in viral proteins, primarily those found in the human viral pathogens.

The predictive approach employed in this study revealed for the first time a large set of putative PrDs in numerous proteins of the emerging human viral pathogens, including those associated with persistent viral infections, oncogenic processes, hemorrhagic fevers, and others. Further analyses of these PrD-containing proteins may improve our understanding of viral infections, and they need to be further expanded along with the discovery of novel viral species under the Virome project^{91–94}.

References

- Eisenberg, D. & Mathias, J. The amyloid state of proteins in human diseases. *Cell* **148**, 6, 1188–1203 (2012).
- Morales, R., Keri, C. & Soto, C. Prion-like features of misfolded A β and tau aggregates. *Virus research* **207**, 106–112 (2015).
- Prusiner, S. Nobel Lecture: Prions. *Proceedings of the National Academy of Sciences* **95**, 13363–13383, <https://doi.org/10.1073/pnas.95.23.13363> (1998).
- Ma, J. Neurotoxicity and Neurodegeneration When PrP Accumulates in the Cytosol. *Science* **298**, 1781–1785, <https://doi.org/10.1126/science.1073725> (2002).
- Stefani, M. Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1739**, 5–25, <https://doi.org/10.1016/j.bbadis.2004.08.004> (2004).
- Shimonaka, S., Nonaka, T., Suzuki, G., Hisanaga, S. & Hasegawa, M. Templated Aggregation of TAR DNA-binding Protein of 43 kDa (TDP-43) by Seeding with TDP-43 Peptide Fibrils. *Journal of Biological Chemistry* **291**, 8896–8907, <https://doi.org/10.1074/jbc.m115.713552> (2016).
- Cobb, N. J. & Witold, K. Prion diseases and their biochemical mechanisms. *Biochemistry* **48**, 12, 2574–2585 (2009).
- Goedert, M., Clavaguera, F. & Tolnay, M. The propagation of prion-like protein inclusions in neurodegenerative diseases. *Trends in Neurosciences* **33**, 317–325, <https://doi.org/10.1016/j.tins.2010.04.003> (2010).
- Furukawa, Y. & Nukina, N. Functional diversity of protein fibrillar aggregates from physiology to RNA granules to neurodegenerative diseases. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1832**, 1271–1278, <https://doi.org/10.1016/j.bbadis.2013.04.011> (2013).
- Prusiner, S. Biology and Genetics of Prions Causing Neurodegeneration. *Annual Review of Genetics* **47**, 601–623, <https://doi.org/10.1146/annurev-genet-110711-155524> (2013).
- Collinge, J. Prion Diseases of Humans and Animals: Their Causes and Molecular Basis. *Annual Review of Neuroscience* **24**, 519–550, <https://doi.org/10.1146/annurev.neuro.24.1.519> (2001).
- Guo, J. & Lee, V. Seeding of Normal Tau by Pathological Tau Conformers Drives Pathogenesis of Alzheimer-like Tangles. *Journal of Biological Chemistry* **286**, 15317–15331, <https://doi.org/10.1074/jbc.m110.209296> (2011).
- Jucker, M. & Walker, L. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature* **501**, 45–51, <https://doi.org/10.1038/nature12481> (2013).
- Shrivastava, S., Steele, R., Ray, R. & Ray, R. MicroRNAs: Role in hepatitis C virus pathogenesis. *Genes & Diseases* **2**, 35–45, <https://doi.org/10.1016/j.gendis.2015.01.001> (2015).
- Shahnawaz, M., Park, K., Mukherjee, A., Diaz-Espinoza, R. & Soto, C. Prion-like characteristics of the bacterial protein Microcin E492. *Scientific Reports* **7**, 45720, <https://doi.org/10.1038/srep45720> (2017).
- Wemheuer, W., Wrede, A. & Schulz-Schaeffer, W. Types and Strains: Their Essential Role in Understanding Protein Aggregation in Neurodegenerative Diseases. *Frontiers in Aging Neuroscience* **9**, <https://doi.org/10.3389/fnagi.2017.00187> (2017).
- Fowler, D., Koulov, A., Balch, W. & Kelly, J. Functional amyloid – from bacteria to humans. *Trends in Biochemical Sciences* **32**, 217–224, <https://doi.org/10.1016/j.tibs.2007.03.003> (2007).
- Sanchez de Groot, N., Torrent, M., Villar-Piqué, A., Lang, B., Ventura, S., Gsponer, J. & Babu, M. Evolutionary selection for protein aggregation. *Biochemical Society Transactions* **40**, 1032–1037, <https://doi.org/10.1042/bst20120160> (2012).
- Espinosa Angarica, V., Ventura, S. & Sancho, J. Discovering putative prion sequences in complete proteomes using probabilistic representations of Q/N-rich domains. *BMC Genomics* **14**, 316, <https://doi.org/10.1186/1471-2164-14-316> (2013).
- Malinowska, L., Kroschwald, S. & Alberti, S. Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1834**, 918–931, <https://doi.org/10.1016/j.bbapap.2013.01.003> (2013).
- Sabate, R. When amyloids become prions. *Prion* **8**, <https://doi.org/10.4161/pri.29238> (2014).
- True, H. & Lindquist, S. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407**, 477–483, <https://doi.org/10.1038/35035005> (2000).

23. Blanco, L., Evans, M., Smith, D., Badtke, M. & Chapman, M. Diversity, biogenesis and function of microbial amyloids. *Trends in Microbiology* **20**, 66–73, <https://doi.org/10.1016/j.tim.2011.11.005> (2012).
24. Chen, S. *et al.* Exposure to the Functional Bacterial Amyloid Protein Curli Enhances Alpha-Synuclein Aggregation in Aged Fischer 344 Rats and *Caenorhabditis elegans*. *Scientific Reports* **6**, <https://doi.org/10.1038/srep34477> (2016).
25. Yuan, A. H. & Hochschild, A. A bacterial global regulator forms a prion. *Science* **355**, 198–201 (2017).
26. Toombs, J., McCarty, B. & Ross, E. Compositional Determinants of Prion Formation in Yeast. *Molecular and Cellular Biology* **30**, 319–332, <https://doi.org/10.1128/mcb.01140-09> (2009).
27. Michelitsch, M. & Weissman, J. A census of glutamine/asparagine-rich regions: Implications for their conserved function and the prediction of novel prions. *Proceedings of the National Academy of Sciences* **97**, 11910–11915, <https://doi.org/10.1073/pnas.97.22.11910> (2000).
28. Prilusky, J. *et al.* FoldIndex(C): a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435–3438, <https://doi.org/10.1093/bioinformatics/bti537> (2005).
29. Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. *Cell* **137**, 146–158, <https://doi.org/10.1016/j.cell.2009.02.044> (2009).
30. Lancaster, A., Nutter-Upham, A., Lindquist, S. & King, O. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* **30**, 2501–2502, <https://doi.org/10.1093/bioinformatics/btu310> (2014).
31. Batlle, C., Iglesias, V., Navarro, S. & Ventura, S. Prion-like proteins and their computational identification in proteomes. *Expert Review of Proteomics* **14**, 335–350, <https://doi.org/10.1080/14789450.2017.1304214> (2017).
32. Iglesias, V., de Groot, N. & Ventura, S. Computational analysis of candidate prion-like proteins in bacteria and their role. *Frontiers in Microbiology* **6**: <https://doi.org/10.3389/fmicb.2015.01123> (2015).
33. March, Z. M., King, O. D. & Shorter, J. Prion-like domains as epigenetic Regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Res. Mar* **17**, <https://doi.org/10.1016/j.brainres.2016.02.037> (2016).
34. Edwards, R. & Rohwer, F. Opinion: Viral metagenomics. *Nature Reviews Microbiology* **3**, 504–510, <https://doi.org/10.1038/nrmicro1163> (2005).
35. Ross, E., MacLea, K., Anderson, C. & Ben-Hur, A. A Bioinformatics Method for Identifying Q/N-Rich Prion-Like Domains in Proteins. *Methods in Molecular Biology* 219–228, https://doi.org/10.1007/978-1-62703-438-8_16 (2013).
36. Toombs, J. *et al.* De novo design of synthetic prion domains. *Proceedings of the National Academy of Sciences* **109**, 6519–6524 (2012).
37. Zambrano, R. *et al.* PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic Acids Research* **43**, W331–W337 (2015).
38. An, L., Fitzpatrick, D. & Harrison, P. Emergence and evolution of yeast prion and prion-like proteins. *BMC Evolutionary Biology* **16** (2016).
39. Tetz G & Tetz V. Prion-Like Domains in Phagobiota. *Frontiers in Microbiology* **8**: <https://doi.org/10.3389/fmicb.2017.02239> (2017).
40. Virgin, H., Wherry, E. & Ahmed, R. Redefining Chronic Viral Infection. *Cell* **138**, 30–50, <https://doi.org/10.1016/j.cell.2009.06.036> (2009).
41. McLaughlin-Drubin, M. & Munger, K. Viruses associated with human cancer. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1782**, 127–150, <https://doi.org/10.1016/j.bbadis.2007.12.005> (2008).
42. Snyderman, D. Characterization of the Reconstructed 1918 Spanish Influenza Pandemic Virus. *Yearbook of Medicine* **2006**, 68–69, [https://doi.org/10.1016/s0084-3873\(08\)70357-2](https://doi.org/10.1016/s0084-3873(08)70357-2) (2006).
43. Sadeghipour, S. & Mathias, R. Herpesviruses hijack host exosomes for viral pathogenesis. *Seminars in Cell & Developmental Biology* **67**, 91–100, <https://doi.org/10.1016/j.semcdb.2017.03.005> (2017).
44. UniProt: a hub for protein information. *Nucleic Acids Research* **43**, D204–D212. <https://doi.org/10.1093/nar/gku989> (2014)
45. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29, <https://doi.org/10.1038/75556> (2000).
46. UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic acids research* **40**, D1, D71–D75, <https://doi.org/10.1093/nar/gkr981> (2011).
47. Adams, M. *et al.* Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses. *Archives of Virology* **162**, 2505–2538, <https://doi.org/10.1007/s00705-017-3358-5> (2017).
48. Colson, P. *et al.* “Megavirales”, a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. *Archives of Virology* **158**, 2517–2521, <https://doi.org/10.1007/s00705-013-1768-6> (2013).
49. De Clercq, E. Strategies In The Design Of Antiviral Drugs. *Nature Reviews Drug Discovery* **1**, 13–25, <https://doi.org/10.1038/nrd703> (2002).
50. Yost, S. & Marcotrigiano, J. Viral precursor polyproteins: keys of regulation from replication to maturation. *Current Opinion in Virology* **3**, 137–142, <https://doi.org/10.1016/j.coviro.2013.03.009> (2013).
51. Bonavia, A., Zelus, B., Wentworth, D., Talbot, P. & Holmes, K. Identification of a Receptor-Binding Domain of the Spike Glycoprotein of Human Coronavirus HCoV-229E. *Journal of Virology* **77**, 2530–2538, <https://doi.org/10.1128/jvi.77.4.2530-2538.2003> (2003).
52. Kobiler, O., Drayman, N., Butin-Israeli, V. & Oppenheim, A. Virus strategies for passing the nuclear envelope barrier. *Nucleus* **3**, 526–539, <https://doi.org/10.4161/nucl.21979> (2012).
53. Gastaldello, S. *et al.* denedylase encoded by Epstein–Barr virus promotes viral DNA replication by regulating the activity of cullin-RING ligases. *Nature Cell Biology* **12**, 351–361, <https://doi.org/10.1038/ncb2035> (2010).
54. Menéndez-Arias, L. & Andino, R. Viral polymerases. *Virus Research* **234**, 1–3, <https://doi.org/10.1016/j.virusres.2017.02.003> (2017).
55. Swanstrom, R. & Wills, J. W. Synthesis, assembly, and processing of viral proteins. *Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY)*, (1997).
56. Chen, D., Jiang, H., Lee, M., Liu, F. & Zhou, Z. Three-Dimensional Visualization of Tegument/Capsid Interactions in the Intact Human Cytomegalovirus. *Virology* **260**, 10–16, <https://doi.org/10.1006/viro.1999.9791> (1999).
57. Chiu, W. & Chang, W. Vaccinia Virus J1R Protein: a Viral Membrane Protein That Is Essential for Virion Morphogenesis. *Journal of Virology* **76**, 9575–9587, <https://doi.org/10.1128/jvi.76.19.9575-9587.2002> (2002).
58. Ostapchuk, P. & Hearing, P. Pseudopackaging of Adenovirus Type 5 Genomes into Capsids Containing the Hexon Proteins of Adenovirus Serotypes B, D, or E. *Journal of Virology* **75**, 45–51, <https://doi.org/10.1128/jvi.75.1.45-51.2001> (2001).
59. Belshaw, R., Pybus, O. & Rambaut, A. The evolution of genome compression and genomic novelty in RNA viruses. *Genome Research* **17**, 1496–1504, <https://doi.org/10.1101/gr.6305707> (2007).
60. Perera, R. & Kuhn, R. Structural proteomics of dengue virus. *Current Opinion in Microbiology* **11**, 369–377, <https://doi.org/10.1016/j.mib.2008.06.004> (2008).
61. Varnum, S. *et al.* Identification of Proteins in Human Cytomegalovirus (HCMV) Particles: the HCMV Proteome. *Journal of Virology* **78**, 13395–13395, <https://doi.org/10.1128/jvi.78.23.13395.2004> (2004).
62. Chan, S., Lee, J., Narula, M. & Ou, J. Suppression of Host Innate Immune Response by Hepatitis C Virus via Induction of Autophagic Degradation of TRAF6. *Journal of Virology* **90**, 10928–10935, <https://doi.org/10.1128/jvi.01365-16> (2016).
63. Benedict, C., Norris, P. & Ware, C. To kill or be killed: viral evasion of apoptosis. *Nature Immunology* **3**, 1013–1018, <https://doi.org/10.1038/ni1102-1013> (2002).
64. Agranovich, A., Maman, Y. & Louzoun, Y. Viral proteome size and CD8+ T cell epitope density are correlated: The effect of complexity on selection. *Infection, Genetics and Evolution* **20**, 71–77, <https://doi.org/10.1016/j.meegid.2013.07.030> (2013).
65. Sugiura, N., Setoyama, Y., Chiba, M., Kimata, K. & Watanabe, H. Baculovirus Envelope Protein ODV-E66 Is a Novel Chondroitinase with Distinct Substrate Specificity. *Journal of Biological Chemistry* **286**, 29026–29034, <https://doi.org/10.1074/jbc.m111.251157> (2011).
66. Kawaguchi, Y., Sugiura, N., Kimata, K., Kimura, M. & Kakuta, Y. The crystal structure of novel chondroitin lyase ODV-E66, a baculovirus envelope protein. *FEBS Letters* **587**, 3943–3948, <https://doi.org/10.1016/j.febslet.2013.10.021> (2013).

67. York, W., Darvill, A., McNeil, M., Stevenson, T. & Albersheim, P. Isolation and characterization of plant cell walls and cell wall components. *Methods in Enzymology* 3–40. [https://doi.org/10.1016/0076-6879\(86\)18062-1](https://doi.org/10.1016/0076-6879(86)18062-1) (1986).
68. Wolf, S., Lucas, W., Deom, C. & Beachy, R. Movement Protein of Tobacco Mosaic Virus Modifies Plasmodesmatal Size Exclusion Limit. *Science* **246**, 377–379. <https://doi.org/10.1126/science.246.4928.377> (1989).
69. Ackermann, H. *Viral Pathogenesis in diagrams*. [S.l.]: CRC press (2017).
70. Kushnirov, V., Vishnevskaya, A., Alexandrov, I. & Ter-Avanesyan, M. Prion and Nonprion Amyloids. *Prion* **1**, 179–184. <https://doi.org/10.4161/prion.1.3.4840> (2007).
71. Hennig, S. *et al.* Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *The Journal of Cell Biology* **210**, 529–539. <https://doi.org/10.1083/jcb.201504117> (2015).
72. Derkatch, I., Bradley, M., Hong, J. & Liebman, S. Prions Affect the Appearance of Other Prions. *Cell* **106**, 171–182. [https://doi.org/10.1016/S0092-8674\(01\)00427-5](https://doi.org/10.1016/S0092-8674(01)00427-5) (2001).
73. Itzhaki, R. Herpes simplex virus type 1 and Alzheimer's disease: increasing evidence for a major role of the virus. *Frontiers in Aging Neuroscience* **6**. <https://doi.org/10.3389/fnagi.2014.00202> (2014).
74. Harris, S. & Harris, E. Herpes Simplex Virus Type 1 and Other Pathogens are Key Causative Factors in Sporadic Alzheimer's Disease. *Journal of Alzheimer's Disease* **48**, 319–353. <https://doi.org/10.3233/jad-142853> (2015).
75. Prevelige, P., Thomas, D. & King, J. Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophysical Journal* **64**, 824–835. [https://doi.org/10.1016/S0006-3495\(93\)81443-7](https://doi.org/10.1016/S0006-3495(93)81443-7) (1993).
76. Guo, L., Han, A., Bates, D., Cao, J. & Chen, L. Crystal structure of a conserved N-terminal domain of histone deacetylase 4 reveals functional insights into glutamine-rich domains. *Proceedings of the National Academy of Sciences* **104**, 4297–4302. <https://doi.org/10.1073/pnas.0608041104> (2007).
77. Reddy, V. & Nemerow, G. Structures and organization of adenovirus cement proteins provide insights into the role of capsid maturation in virus entry and infection. *Proceedings of the National Academy of Sciences* **111**, 11715–11720. <https://doi.org/10.1073/pnas.1408462111> (2014).
78. Rux, J., Kuser, P. & Burnett, R. Structural and Phylogenetic Analysis of Adenovirus Hexons by Use of High-Resolution X-Ray Crystallographic, Molecular Modeling, and Sequence-Based Methods. *Journal of Virology* **77**, 9553–9566. <https://doi.org/10.1128/jvi.77.17.9553-9566.2003> (2003).
79. Ebner, K., Pinsker, W. & Lion, T. Comparative Sequence Analysis of the Hexon Gene in the Entire Spectrum of Human Adenovirus Serotypes: Phylogenetic, Taxonomic, and Clinical Implications. *Journal of Virology* **79**, 12635–12642. <https://doi.org/10.1128/jvi.79.20.12635-12642.2005> (2005).
80. Rux, J. & Burnett, R. Type-Specific Epitope Locations Revealed by X-Ray Crystallographic Study of Adenovirus Type 5 Hexon. *Molecular Therapy* **1**, 18–30. <https://doi.org/10.1006/mthe.1999.0001> (2000).
81. Onion, D. *et al.* The CD4+ T-cell response to adenovirus is focused against conserved residues within the hexon protein. *Journal of General Virology* **88**, 2417–2425. <https://doi.org/10.1099/vir.0.82867-0> (2007).
82. Molliex, A. *et al.* Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **163**, 123–133 (2015).
83. Kato, M. *et al.* Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* **149**, 753–767 (2012).
84. Decker, C. & Parker, R. P-Bodies and Stress Granules: Possible Roles in the Control of Translation and mRNA Degradation. *Cold Spring Harbor Perspectives in Biology* **4**, a012286–a012286 (2012).
85. Batlle, C., de Groot, N., Iglesias, V., Navarro, S. & Ventura, S. Characterization of Soft Amyloid Cores in Human Prion-Like Proteins. *Scientific Reports* **7** (2017).
86. Chuan, Y., Fan, Y., Lua, L. & Middelberg, A. Virus assembly occurs following a pH- or Ca²⁺-triggered switch in the thermodynamic attraction between structural protein capsomeres. *Journal of The Royal Society Interface* **7**, 409–421 (2009).
87. Ding, Y., Chuan, Y., He, L. & Middelberg, A. Modeling the competition between aggregation and self-assembly during virus-like particle processing. *Biotechnology and Bioengineering* **107**, 550–560 (2010).
88. Vernon, R. *et al.* Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* **7** (2018).
89. Fraser, N. & Valyi-Nagy, T. Viral, neuronal and immune factors which may influence herpes simplex virus (HSV) latency and reactivation. *Microbial Pathogenesis* **15**, 83–91. <https://doi.org/10.1006/mpat.1993.1059> (1993).
90. Fuller, E., Elder, B. & Dwyer, G. Pathogen Persistence in the Environment and Insect-Baculovirus Interactions: Disease-Density Thresholds, Epidemic Burnout, and Insect Outbreaks. *The American Naturalist* **179**, E70–E96. <https://doi.org/10.1086/664488> (2012).
91. Carroll, D. *et al.* The Global Virome Project. *Science* **359**, 872–874. <https://doi.org/10.1126/science.aap7463> (2018).
92. Tetz, G. *et al.* Bacteriophages as potential new mammalian pathogens. *Scientific Reports* **7**. <https://doi.org/10.1038/s41598-017-07278-6> (2017).
93. Tetz, G. & Tetz, V. Bacteriophage infections of microbiota can lead to leaky gut in an experimental rodent model. *Gut Pathogens* **8**. <https://doi.org/10.1186/s13099-016-0109-1> (2016).
94. Yi, F. *et al.* Epstein-Barr virus nuclear antigen 3C targets p53 and modulates its transcriptional and apoptotic activities. *Virology* **388**, 236–247. <https://doi.org/10.1016/j.virol.2009.03.027> (2009).

Author Contributions

G.T. and V.T. designed and conducted the experiments and analyzed data. G.T. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-27256-w>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018