Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Nipah shell disorder, modes of infection, and virulence

Gerard Kian-Meng Goh\textsuperscript{a,∗}, A. Keith Dunker\textsuperscript{b}, James A. Foster\textsuperscript{c,d}, Vladimir N. Uversky\textsuperscript{e,f}

\textsuperscript{a} Goh’s BioComputing, Singapore, Republic of Singapore
\textsuperscript{b} Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN, USA
\textsuperscript{c} Department of Biological Sciences, University of Idaho, Moscow, ID, USA
\textsuperscript{d} Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID, USA
\textsuperscript{e} Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL, USA
\textsuperscript{f} Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow region, Russia

A R T I C L E   I N F O

Keywords:
Intrinsically disordered protein
Nucleocapsid
Nipah
Virulence
Viral protein
Protein structure
Protein function
Shell

A B S T R A C T

The Nipah Virus (NiV) was first isolated during a 1998–9 outbreak in Malaysia. The outbreak initially infected farm pigs and then moved to humans from pigs with a case-fatality rate (CFR) of about 40%. After 2001, regular outbreaks occurred with higher CFRs (~71%, 2001–5, ~93%, 2008–12). The spread arose from drinking virus-laden palm date sap and human-to-human transmission. Intrinsic disorder analysis revealed strong correlation between the percentage of disorder in the N protein and CFR (Regression: \( r^2 = 0.93, p < 0.01 \)). Distinct disorder and, therefore, genetic differences can be found in all three group of strains. The fact that the transmission modes of the Malaysia strain are different from those of the Bangladesh strains suggests that the correlations may also be linked to the modes of viral transmission. Analysis of the NiV and related viruses suggests links between modes of transmission and disorder of not just the N protein but, also, of M shell protein. The links among shell disorder, transmission modes, and virulence suggest mechanisms by which viruses are attenuated as they passed through different cell hosts from different animal species. These have implications for development of vaccines and epidemiological molecular analytical tools to contain outbreaks.

1. Introduction

The Nipah virus (NiV) was first discovered, when an outbreak happened in Malaysia and Singapore in 1998–99. About 270 people were infected with a case-fatality ratio (CFR) of 40% [1–3]. All of the cases involved close contacts with infected farm pigs. Many of the pigs were initially infected by eating fruits that were partially eaten by fruit bats, which had left their virus-laden saliva behind (4–6). Upon isolation and sequencing of the virus, it was revealed that the culprit was a novel RNA virus from \textit{Paramyxoviridae} family, which includes the Hendra (HeV), mumps (MuV), and measles (MeV) viruses [4–7]. The symptoms for NiV infections often include an initial flu-like fever that often follows by irritations, coma, and then death.

In 2001, a different NiV strain was observed in Bangladesh [4]. There were outbreaks in India or Bangladesh each year thereafter. While human infections in the Malaysian outbreak arose mainly from close contacts with infected pigs, the patients in Bangladesh were often infected by drinking date palm sap that was previously consumed by bats. The virulence was also different. For example, the Bangladesh outbreaks in 2004 involved a CFR of 75%, while the 2008–10 outbreaks had a CFR of 89% (Kalkarni et al. [8]). The reason for these remarkable differences in the virulence of different NiV strains remains largely unknown [9]. This paper will put forth a set of evidence demonstrating that the variability in the intrinsic disorder propensities of some of the viral proteins from different strains NiV may act as an underlying reason for the differences in their virulence with the potential links to the modes of transmission. It is worth mentioning here the same underlying reason for high virulence and differences in the transmission modes had been previously found for other viruses, such as MERS-CoV, SARS-CoV, Ebola virus, flaviviruses, and HIV [10–13,67].

An important concept that will be used in this paper involves intrinsically disordered proteins, which are biologically active proteins that have no unique structures. While it had been initially observed that unique structures dictate the functions of the protein, it was also found that many proteins lack structure, and the lack of structure itself provides protein with new functional means [14–17]. As a result of the recognition of the biological and pathological importance of intrinsically disordered proteins and proteins containing intrinsically
disordered regions, multiple disorder predictors have been elaborated to recognize disordered regions [18–20]. One of the earliest disorder predictors is the PONDR® VLXT [21–23] which has been shown to be useful in the analysis of proteins from various viruses, such as HIV, influenza A 1918H1N1 and H5N1 viruses, poliovirus, SARS-CoV, MERS-CoV, smallpox virus, Ebola virus, and HCV [10–13,24–31,67].

While NiV is of the Paramyxoviridae, this virus, together with HeV, belongs to the henipavirus genus characterized by larger genomes, when compared to the other paramyxoviruses. In fact, 18.6 kb-long negative-sense single-stranded RNA (ssRNA) genome of NiV contains 6 genes (Fig. 1) that encode 9 proteins: nucleoprotein (N), phosphoprotein (P), the interferon antagonists W and V, the viral C protein, a matrix protein (M), viral fusion and glycoproteins (F and G, respectively), and a large polymerase (L), with the proteins P, V, W, and C about the genomic organization, which are used by the virus to overcome the innate immune response [32,33], being all encoded by the P gene that undergoes specific mRNA editing that results in the appearance of the specific reading frame shifts [34,35]. A homopolymer of the major nucleocapsid protein N condenses and encases the genomic RNA within the long helical nucleocapsids, which serve as biologically active templates for the viral RNA synthesis by the viral RNA-dependent RNA polymerase [25,35,36]. Structurally, N protein contains two globular domains, the N-terminal (NNTD, residues 32–258) and C-terminal (NCCTD, residues 259–371) domains decorated with two projections, N-terminal (NTARM, residues 1–31) and C-terminal subdomains (CTARM, residues 372–383), and a long and highly disordered C-terminal tail (NTAIL, residues 384–532) that protrudes outside the nucleocapsid. These different structural parts of N have divergent biological functions, where NNTD and NCCTD enwrap the genomic RNA to protect it against nucleases [37–41], NTARM and CTARM from adjacent protomers are exchanged to ensure stable lateral contacts needed for stabilization of the N homopolymer 2005) [37–41], and a highly disordered Ntail is utilized in binding to the C-terminal domain of the phosphoprotein P (PXD) [42–54]. Therefore, the viral shell contains N as a major nucleocapsid protein alongside with the phosphoprotein, P, and a large protein L that serves as an RNA polymerase during viral replication [6,8,52,55], and which is also found in close proximity. The matrix protein, M, that can be found underlying the viral membrane, is in contact with the fusion protein, F, and the glycoprotein, G, bound to the lipid membrane [7].

2. Material and methods

PONDR® VLXT (www.pondr.com) is the main tool used in our investigation. It is a neural network that reads inputs from the sequence of a protein and then predicts regions that are expected to be disordered as output [21,23]. The PONDR® VLXT has been found to be the most accurate predictor in situations that involve protein-protein interactions, which are essential in viral shell proteins [56, 57]. This is likely the reason for it to be rather successful in the various studies of viral shell proteins as aforementioned.

The sequences of the various NiV strains were selected and downloaded from UniProt (http://www.uniprot.org). The programming language JAVA was then used to retrieve and record into MYSQL. This process was done through a JAVA-JDBC driver. The program also calculates the percentage of intrinsic disorder (PID), which is defined as the number of predicted disordered residues divided by the total number of residues in the protein. In order to obtain a clearer visualization, a search for Nipah nucleocapsid was made using NCBI-Protein Databank (NCBI-PDB, http://www.ncbi.nlm.nih.gov/Structure/index.shtml). The accession codes were recorded and the respective PDB and FASTA files were downloaded and placed into the MySQL database. The information stored in the database can later be retrieved using a JAVA program that would generate codes to be used in molecular viewing software, Jmol (www.jmol.org) [29]. Statistical analyses were done using one-way analysis of variance (ANOVA) via R statistical package. The case-fatality ratios were calculated based on published data from the World Health Organization (WHO) (http://www.searo.who.int/entity/emerging_diseases/links/nipah_virus_outbreaks_sear/en/).

A phylogenetic tree was constructed using Clustal Omega (EMBL-EBI, https://www.ebi.ac.uk/Tools/msa/clustalo/) and the mentioned sequences of NiV N proteins from UniProt. The NiV strains and the respective calculated N PIDs (disorder levels) were carefully denoted on the tree.

3. Results

3.1. Nipah virulence and CFR analysis by year of infection

It has already been known that the virulence of the Malaysia-Singapore 1998–9 NiV strain was lower than those of strains that appeared in Bangladesh and India in later years. This can be reaffirmed by looking at Fig. 2 showing a CFR plot of the historical records from the WHO. It should be noted that the plot is adjusted for the “noise”. The CFR of the 1998–9 outbreak, which involved 270 people, was about 40%, while subsequent 2004 outbreaks in Bangladesh and India showed CFR of ~75%, and the outbreaks caused by the NiV strains after 2007 showed even higher fatality rates, which average at about 85%. Statistical analyses using One-Way ANOVA (F-9.93, p < 0.01) and regression reaffirm the statistical significance of the differences in the virulence strength among the strains (r² = 0.83, F = 42.43, p < 0.01). These results therefore justify the grouping of NiV into three main strains with the following calculated averages: 1) Malaysia-Singapore 1998–9 (CFR, 39.5 ± 1%), 2) Bangladesh-India 2001–5 (CFR, 71 ± 3.5%), and 3) Bangladesh-India 2007/8–12 (92.8 ± 6%). As we shall see from the analyses below, there is also noticeable difference in the viral shell disorder and, therefore, genetic difference, among these three groups of the NiV strains.
the PONDR insights into the reasons for the observed PID di-

sequences. Fig. 3A represents the PONDR of di-

Fig. 3 shows that mutations are causing noticeable distortions in the local intrinsic disorder predisposition and that these distortions typically propagate well beyond the mutation site. Furthermore, Fig. 3 illustrates that although N proteins generally contain high levels of intrinsic disorder, their C-terminal domains (residues 370–532) are almost completely disordered. This is a peculiar observation indicating that strain-specific mutations are predominantly located within these highly disordered C-terminal domains suggesting their importance for the strain evolution, probably via modulation of the biological activities of the Nipah N protein.

3.4. Strain-specific differences tend to be located near the C-terminus of the N protein used for interactions with the P protein

As it follows from Fig. 3, most strain-specific differences are found within the highly disordered C-terminal regions of the N proteins. This is a very important observation, since previous structural analysis revealed that this region represents an interaction site of N nucleoprotein for the P protein, and the binding of N to P is crucial for the appropriate co-localization of the complex during viral replication [52,59]. Fig. 4 shows that the C-tail of the N protein contains several disorder-based binding sites, molecular recognition features (MoRFs), which can undergo disorder-to-order transitions at interaction with their biological partners, and which can be identified by the ANCHOR algorithm [60,61]. Fig. 4 also illustrates that the strain-specific mutations are expected to affect interactability of N-protein, since all of them increase ANCHOR scores of the C-terminal region.

The crystal structure of the N–P protein complex with disorder annotations in N protein shown in Fig. 5 re-affirms the mostly C-terminal location of the regions with strain-specific differences in disorder (pink) among strains. The figure is somewhat misleading by showing only one pink region (around residue 315). This is due to the fact that the other regions with strain-specific variability in disorder content shown in Fig. 3 are missing because of the fact that the construct used in the crystallization experiments (residues 28–383) exclusively contained the globular NNTD and NCTD domains and CTARM, whereas NTARM and NTail were removed to avoid complications associated with an attempt to crystallize long regions with high disorder content and to avoid

| Outbreak Strains | Number of Cases (Deaths) | Fatality Rate (CFR, %)* | PID (%) N | UniProt accession numbers |
|------------------|--------------------------|-------------------------|-----------|--------------------------|
| Malaysia-Singapore 1998-99 | 276 (106) | 38 | 41.73 ± 0.01b | Q9K092, A0A1LK92 |
| Bangladesh-India 2001–2005 | 170 (123) | 72 | 42.67 ± 0.27b | Q4VQ2, H6V858 |
| Bangladesh-India 2007-10 | 103 (88) | 85 | 43.32 ± 0.24b | D2DEB7, A0A141SU6, H6V855, H6V857, E4Z9F9d |

Statistical Analyses: ANOVA, p < 0.01, Regression Analysis, p < 0.01, r² = 0.93.

* CFR are calculated from the total number of death divided by the total number of cases.

+ The numeric value after ‘+’ is the standard error. It should be noted that the mean PID values are misleadingly close to each other. It does not need to be the case since the N protein is huge and most of the mutations arise from or from areas close to the Ntail.

+ Samples from two anamalous strains are not included: H6V868, H6V854. In general, a few anamalous strains should be expected in cases where a large number of outbreaks/strains is involved. The data used involve 9 isolates and 6 strains.

+ This Thailand sample was collected in 2010 from bats. This sample makes the study interesting as the PID and genetic makeup is closer to those of the Bangladesh-India 2007-10 group despite the fact that Thailand is geographically closer to Malaysia-Singapore. However, because the PID and genetic makeup is closer to that of the 2008-10 strains and because 2007 is close to 2008, we suspect that this strain falls into the 2008-10 strains.

+ Sample was collect from Malaysian bats in 2008. It is of great interest since its PID and genetic makeup is closer to the Bangladesh-India 2007-10 samples than to those of Malaysia-Singapore 1998–99.
unwanted self-assembly of N \[62\]. Interestingly, Fig. 5 shows the pink area virtually touching the protein, P, which is shaded in dark blue. Although the most ordered part of N protein was used in the crystalization experiments and although N protein was co-crystallized with the P nucleocapsid-binding domain PXD (residues 655–709), the target complex formed a trimer of heterodimers, where all involved components contained significant amount of intrinsic disorder, as can be judged by the presence of noticeable regions of missing electron density. In fact, all three chains of N found in this trimer of heterodimers contained variable levels of structural disorder, with residues 28–31, 116–121, 186–189, and 372–383 in chain A, residues 28–31, 118, 182–187, 241–247, 346–347, and 370–383 in chain B, and residues 28–42, 61–64, 79–153, 169–199, 239–248, and 369–383 in chain C being located within regions with missing electron density. Similarly, 12 to 14 of the C-terminal residues of the three chains of the PXD were missing in the trimer of heterodimers structure (PDB ID: 4CO6) \[62\].

3.5. Nipah phylogenetic and disorder analyses

Phylogenetic analyses of NiV have already been performed in the past (1,4). For instance, Lo Presti et al. used the NiV N protein in the construction of a phylogenetic tree that suggests that NiV first entered
Southeast Asia around 1947. We have made a phylogenetic analysis to compliment our disorder study as seen in Fig. 6. Fig. 6 is able to reveal results that would have been impossible to obtain using disorder or phylogenetic method alone. We are able to see that the Malaysia strains cluster together with strains from nearby regions such as Thailand regardless of differences in both N disorder (PID) and year. This reasserts past observations of co-evolutions among strains from different regions i.e. Bangladesh-India and Malaysia [1].

4. Discussion

4.1. Phylogenetic and disorder evidence of co-evolutions and possible roles of migratory bats

We have demonstrated the presence of a correlation between the Nipah virulence and its N protein disorder. The virulence of the Nipah virus can be traced to the year that the virus sample was obtained even more so than the geographic localization from which the sample was retrieved. For example, samples taken from Malaysia and Thailand in 2008 and 2010 have PID values closer to those of samples from Bangladesh-India 2007-12 than to those of Bangladesh-India 2001-5 and Malaysia-Singapore 1998-9, despite the fact Malaysia is in close proximity to Thailand. Without phylogenetic analysis, it would be easy to misinterpret the results by assuming that migratory bats may have brought the various strains of different PID from elsewhere. Our combined disorder and phylogenetic analysis says, however, that this is not necessarily the case. As seen in Fig. 6, the Thailand 2010, Malaysia 2010 and 1998 strains are all clustered together despite huge differences in year and PID. Thailand and Malaysia are in close geographic proximity together. This basically implies that the strains co-evolved at distinct locations (i.e. Southeast Asia and Bangladesh) as already suggested by others [1,4], even if the various strains in separate locations assume different N disorder as seen in Fig. 6.

Strangely, though, the phylogenetic-disorder analysis also reveals that there is room for the role of migratory bats as Fig. 6 shows that a specific strain from Bangladesh-2008 (H6V855) is closer to the strains from Southeast Asia than many of its counterparts from Bangladesh-India, regardless of year and PID. This last observation suggests that there is some circulation of strains over the long distances, presumably with the help of migratory bats. A previous study, using GPS, supports this hypothesis by showing that bats can travel a round trip of over 1000 km in just one month [63]. While our study presents some cases of crossover strains from migratory bats, it also conversely shows also the strong presence of co-evolution as explained above. This study therefore paints a picture of the virus strains facing various but often similar evolutionary pressures at distinct geographic regions. The pressures could have arisen from factors which are the results of varying
preferential modes of transmission dependent on the species of bat hosts and other host types such as pigs and humans. The pressures are therefore evolutionarily manifested as the differences in N PID as we shall re-visit later.

4.2. C-terminus of N protein touches P protein: the focal point

The N, P, and L proteins play important roles in the replication of the viral RNA [64,65]. The N protein when bound to the P protein, encapsidates the viral RNA during the replication of the latter. When P binds to N, the complex stabilizes in order to have greater affinity for the viral RNA. Also, when P protein binds to N protein, it induces the binding of the L protein close to the binding region of the N protein, and together the N-P-L complex functions as an RNA polymerase during the viral replication [52,64].

The fact that most of the areas affected by strain-specific mutations lie near the C-terminus of N protein (Figs. 3-5), which represent or are located in close proximity to the regions that bind to the P protein (Fig. 5), suggests that the greater disorder of the more virulent strains may be involved in modulating the aforementioned functions of the viral proteins N, P, and L. It is likely that the regions of greater disorder found in the more virulent strains, such as that shown in pink in Fig. 5, help to recognize and assist in the binding of N to P and perhaps, also to L. A more efficient replication process will mean greater viral load that makes the strain more virulent.

4.3. Virulence arises from "Trojan horse" replication: reproducibility of past research

We need to be reminded that correlations between the inner shell and the virulence of the virus have been determined not just in Nipah virus as we have shown in the current study, but also in Ebola virus and flaviviruses, such as dengue and yellow fever viruses [10,11,30,67]. This is a reflection of the reliability of the applications of PONDR® VLXT for the analysis of viral shell proteins and is based on the fact that viral shells often have similar functions across the different viral species. While high disorder at the outer shell represents a way of immune evasion that involves the inability to neutralize the virus easily, greater disorder at the inner shell represents a different means of the immune evasion. The former has been found to be a “true viral shapeshifters” that includes the HIV, whereas the latter class has been describe as “Trojan horses”, which are the viral shapeshifters of a different kind. It is named a “Trojan horse” as its strategy to evade the host immune system involves rapid reproduction before the immune system can recognize the foreign invader.

4.4. Empirical evidence of a correlation between the disorder status of N protein and NIV virulence in the wild

Interestingly, past experiments using mutations of N proteins of NIV cousins, canine distemper virus (CDV) and MeV, have shown that greater disorder at the Ntail cause the viruses to be more virulent to ferrets in a laboratory [64]. While this was done experimentally using MeV and CDV in a laboratory setting, our results show for the first time that this also can be the case for NIV, but using an empirical analysis of NIV in the wild. An additional interesting note is the fact that mutations that affect the pathogenesis of the virus is found further upstream to location 315 as illustrated by Figs. 3–5, not noticed by any previous literature. Apparently, greater disorder around 315 facilitates easier binding of the P protein to the C-terminus of N protein.

The findings above are a reflection of the way the virus evolves. While they obviously do not provide all the answers, these observations do provide hints of how we could go about seeking answers to some of the questions involving Nipah evolution. We know that the 1998-9 human outbreak in Malaysia-Singapore involved mainly transmission by close contacts with infected pigs. The outbreaks in Bangladesh-India, however, involved patients consuming virus-laden date palm sap or human-to-human transmission [1,2]. Little is currently known about bat-to-bat transmission in the wild. It is likely that consideration of the variability in PIDs of N proteins could also provide an evolutionary clue on Nipah transmission and its intricate link to virulence.

4.5. Evidence of a link between NIV transmission modes, virulence, and shell disorder

The fact that Nipah strain from 1998 to 9 has a PID value for N protein of 41.73 ± 0.01%, whereas those of N proteins from the 2001–5 and 2007-12 strains are increasing to 42.67 ± 0.27% and 43.32 ± 0.24%, suggests that the differences in the disorder levels of N protein of the Malaysia-Singapore and Bangladesh-India strains can be associated with the differences in the preferred modes of transmission of NIV. Such findings are in line with the results of previous studies where it has been shown that changes in the disorder of the shells can be used to predict the mode of transmission of coronaviruses, including SARS-CoV and MERS-CoV [25,26]. In those studies, we have found that the more disordered the shells of coronaviruses are, the greater the contribution of the respiratory transmission and the lesser the fecal-oral transmission component. The results of the current research seem to extend this finding from coronaviruses to NIV (and, probably, to other RNA viruses from Paramyxoviridae family). In fact, we can see that the lower PID level of N protein from the 1998-9 strain allows greater fecal-oral component, such that contacts with fecal and bodily fluid of pigs provide for easier spread of the virus to the farmers. The higher PIDs of N protein in the 2001-12 strains, on the other hand, facilitate greater chances of human-to-human transmission by respiratory contacts. Further evidence of links between the shell disorder and modes of transmission can also be found in other viruses, including HIV and other retroviruses.

It should be noted that we had previously been able to establish a more comprehensive relationship between coronavirus shell disorder and modes of transmission because much was already known about the behaviors of various porcine coronaviruses. On the contrary, little is currently known about modes of NIV transmission among bats in the wild. The results of the analysis reported in this paper suggest the ways we can further study NIV in both the wild and laboratories using disorder levels of N protein as a useful proxy. Related to this, the results reported in this paper highlight one reason for the differences among the virulence of various NIV strains is likely to be arising from the evolutionary pressures pertaining to the fitness of a particular mode of transmission. For example, a bat-to-bat transmission may favor respiratory transmission, whereas bat-to-pig and pig-to-human transmissions could favor the fecal-oral route. Fig. 7 re-affirms that this link is not confined to NIV but extends to some of the NIV’s cousins with respect to the disorder levels of not just N protein but also that of the M.
contrast, MeV spreads virtually entirely by respiratory mode [7,66]. This is consistent with the tendency of the N and M proteins of the N and M proteins of the various NiV cousins. For comparative purposes, only maximal PIDs found for a virus and protein are used.

4.6. Links between shell disorder and modes of viral transmission across the cousins

A look at Fig. 7 allows us to see the link between the viral shell disorder and the modes of viral transmission, especially across the NiV relatives. The viral shells include both N and M proteins. A noticeable trend is the increase in the disorder of the shell proteins as the components of the respiratory and fecal-oral modes increases and decrease respectively. This is consistent with findings with regard to other viruses in the past. Viruses that rely heavily on fecal-oral transmission tend to have harder shells that are needed to persist in the environment for a longer period of time. On the other hand, viruses that rely on respiratory transmission usually require close contacts that involve bodily fluids (6). NiV spread is more complicated. While NiV transmission from bats to humans or pigs usually involves consumption of fecal or salivary contaminated fruits or palm date saps, the human transmission of NiV Malaysia-Singapore 1998-9 strain involved only close human contacts with infected pigs via bodily fluids, unlike the evidence of Bangladesh-India human-to-human transmission, which is usually indicative of the presence of the respiratory mode [10,25,26]. In sharp contrast, MeV spreads virtually entirely by respiratory mode [7,66].

A look at Fig. 7 allows us to see the link between the viral shell disorder and the modes of viral transmission, especially across the NiV relatives. The viral shells include both N and M proteins. A noticeable trend is the increase in the disorder of the shell proteins as the components of the respiratory and fecal-oral modes increases and decrease respectively. This is consistent with findings with regard to other viruses in the past. Viruses that rely heavily on fecal-oral transmission tend to have harder shells that are needed to persist in the environment for a longer period of time. On the other hand, viruses that depend on respiratory modes of transmission have moderate disorder on their outer shells, but are likely to have highly disordered inner shells. Sexually transmitted viruses, such as HIV-1, on the other hand, could have highly disordered outer shell and have inner shells with reasonably high disorder levels. These are the trends we are seeing in Fig. 7.

This analysis can be further extended by considering a correlation of the disorder status N proteins from various NiV and the transmission modes of these strains. The Bangladesh NiV strains that are characterized by higher respiratory transmission component have highest PID values. Furthermore, while the maximal PID value determined among the N proteins of the various NiV strains is higher than that of the HeV N PID, a comparison of Table 1 and Fig. 7 shows that the N protein from the Malaysia-Singapore NiV strain has an PID value (41.73 ± 0.01%) comparable to that of the N protein from HeV (PID: 41%). In agreement with their lower PID values, both the Malaysia-Singapore NiV and HeV have greater fecal-oral components, since the transmissions of both viruses require close contacts in the presence of bodily fluids. This refers to the horse-to-horse and pig-to-human transmissions of the HeV and Malaysia-Singapore 1998-9 NiV strain, respectively. No human-to-human spread was observed in this NiV strain.

The MeV, however, represents an interesting contrast, as, unlike HeV and NiV, it relies mostly on respiratory transmission for its spread. Fig. 7 shows that PIDs of both M and N proteins from MeV are noticeably higher than those of its somewhat distant cousins shown. The maximal PID values of N proteins from HeV and NiV are 41% and 43% respectively, whereas that of the MeV is 46%. Similarly, the maximal PID values for MeV, HeV, and NiV are 20%, 8%, and 14%, respectively. These observations suggest that the evolutionary need of persistence in the environment (which is associated with the disorder levels of the viral N and M proteins) depends on the modes of transmission that the viruses rely on.

HIV-1 is, of course, an unrelated virus, but it makes an interesting reference point in Fig. 7, as it is sexually transmitted. It has been theorized that the highly disordered outer shell of HIV-1 makes it difficult for an effective vaccine to be found. As we can see from Fig. 7, none of the PID values for the M proteins of the paramyxoviruses are higher than that of HIV. Therefore, the successful development effective vaccines is likely to be feasible for NiV, MeV, and HeV. In agreement with this hypothesis, the vaccines for MeV and HeV have been already elaborated [6,7].

5. Conclusion

There have been few studies on the virulence and infectivity of NiV. The results of this paper correlate infectivity and virulence of NiV to the levels of intrinsic disorder in the N protein and hints to the relationship of disorder levels to modes of viral transmission. These findings have multiple implications for the future research into not just NiV virulence and infectivity, but also that of related viruses, as the same protein is likely to be playing the same role among related viruses. Needless to say, the Paramyxoviridae family is an important class of viruses that includes some medically interesting ones, such as MeV, CDV, HeV and, of course, NiV.

It should be understood that the differences in disorder of the N protein and the behavior of the virus in terms of virulence and infectivity can be related to the way the virus evolves. We have seen that virulence in NiV is the result of the virus’s “Trojan horse” strategy of producing viral particles rapidly before the host immune system is able to detect and neutralize them. The greater virulence is likely to have arisen from the higher viral load of strains that comes with greater N protein disorder.

Given that the level of N protein disorder is associated with NiV virulence, this research may pave a way for some new strategies for the development of attenuated vaccines, not just against HIV, but also against viruses related and unrelated to NiV. Another possible application is the potential of disorder analysis of N protein disorder as a tool for analyzing the outbreaks and deciding policies and approaches for preventing further spread. The detection of links between modes of transmission and shell disorder will also be particularly helpful in the development of tools to anticipate the nature of an outbreak via molecular analysis of the viral shells. Since little is currently known about NiV virulence and infectivity, any new knowledge pertaining to the potential virulence and infectivity of a virus in an outbreak can prove to be invaluable.

Lastly, the results reported here suggest an answer to an age-old biomedical mystery, where it is not understood why viruses become attenuated, when they are passed to the cells that are from different regions.
animal species and of different cell types for many generations as in the case of the preparation of Sabin polio vaccine [66]. Our results seem to hint that when the host cell is changed, there will be a switch of preference for a different mode of transmission as the cell type and animal species have been changed, and the in process the virus becomes attenuated. The “preference” is likely the result of the way the virus evolves or has evolved to optimally adjust to the cell types and the animal species.

Funding

JAF is funded in part by NSF DBI0939454, BEACON Center for Evolution in Action.

Declaration of competing interest

GKM is an independent researcher, and the owner of Goh’s BioComputing, Singapore. GKM has also written a book on a related subject. The authors have no other potential conflict of interests.

References

[1] S. Angeletti, A. Lo Presti, E. Celli, M. Ciccozzi, Molecular epidemiology and phylogeny of nipah virus infection: a mini review, Asian Pac J Trop Med 6 (2013) 1286–1292.
[2] B.S.P. Ang, T.C.C. Lim, L. Wang, Nipah virus infection, J. Clin. Microbiol. 56 (2018) e1875–17.
[3] Mire et al. (2018), C.E. Mire, B.A. Satterfield, J.B. Geisbert, K.N. Agans, V. Borisevich, L. Yan, Y.P. Chan, R.W. Cross, K.A. Fenton, C.C. Broder, et al., Pathogenic differences between nipah virus Bangladesh and Malaysia strains in primates: implications for antibody therapy, Sci. Rep. 6 (2016) 30916.
[4] A. Lo Presti, E. Cella, A. Lai, S. Angeletti, G. Zehender, M. Ciccozzi, G.K.-M. Goh, et al., Microbial Pathogenesis 141 (2020) 103976.
[5] B.S.P. Ang, T.C.C. Lim, L. Wang, Nipah virus infection, J. Clin. Microbiol. 56 (2018) e1875–17.
[6] Mire et al. (2018), C.E. Mire, B.A. Satterfield, J.B. Geisbert, K.N. Agans, V. Borisevich, L. Yan, Y.P. Chan, R.W. Cross, K.A. Fenton, C.C. Broder, et al., Pathogenic differences between nipah virus Bangladesh and Malaysia strains in primates: implications for antibody therapy, Sci. Rep. 6 (2016) 30916.
[7] A. Lo Presti, E. Cella, A. Lai, S. Angeletti, G. Zehender, M. Ciccozzi, G.K.-M. Goh, et al., Microbial Pathogenesis 141 (2020) 103976.
[8] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[9] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[10] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[11] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[12] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[13] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[14] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[15] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[16] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[17] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[18] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[19] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[20] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[21] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[22] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[23] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[24] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[25] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[26] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[27] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[28] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[29] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[30] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[31] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[32] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[33] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[34] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[35] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[36] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[37] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
[38] E. de Witt de Wit, V.J. Munster, Animal models of disease shed light on nipah virus pathogenesis, Virology 364 (2007) 1–14.
alignments, Biochemistry 46 (2007) 13468–13477.

[57] C.J. Oldfield, Y. Cheng, M.S. Cortese, P. Romero, V.N. Uversky, A.K. Dunker, Coupled folding and binding with alpha-helix-forming molecular recognition elements, Biochemistry 44 (2005) 12454–12470.

[58] K. Rajagopalan, P. Mooney, N. Parekh, R.H. Getzeng, A majority of the cancer/testis antigens are intrinsically disordered proteins, J. Cell. Biochem. 112 (2011) 3256–3267.

[59] M. Omi-Furutani, M. Yoneda, K. Fujita, F. Ikeda, C. Kai, Novel phosphoprotein-interacting region in nipah virus nucleocapsid protein and its involvement in viral replication, J. Virol. 84, (2010) 9793–9799.

[60] B. Meszaros, I. Simon, Z. Dosztanyi, Prediction of protein binding regions in disordered proteins, PLoS Comput. Biol. 5 (2009) e1000376.

[61] Z. Dosztanyi, B. Meszaros, I. Simon, Anchor: web server for predicting protein binding regions in disordered proteins, Bioinformatics 25 (2009) 2745–2746.

[62] F. Yabukarski, P. Lawrence, N. Tarbouriech, J.M. Bourhis, E. Delaforge, M.R. Jensen, R.W. Rujgrok, M. Blackledge, V. Voelkhov, M. Jamin, Structure of nipah virus unassembled nucleoprotein in complex with its viral chaperone, Nat. Struct. Mol. Biol. 21 (2014) 754–759.

[63] T.J. Weller, K.T. Castle, F. Liechti, C.D. Hein, M.R. Schirmacher, P.M. Cryan, First direct evidence of long-distance seasonal movements and hibernation in a migratory bat, Sci. Rep. 6 (2016) 34585.

[64] V.D. Thakkar, R.M. Cox, B. Sawatzky, R. da Fontoura Budszewski, J. Sourimant, K. Wabbel, N. Mahsouli, A.L. Greninger, V. von Messling, R.K. Piemper, The unstructured paramyxovirus nucleocapsid protein tail domain modulates viral pathogenesis through regulation of transcriptase activity, J. Virol. 92 (2018).

[65] S.T. Ong, K. Yunos, C.L. Kho, J.O. Abdullah, W.S. Tan, Mutagenesis of the nucleocapsid protein of Nipah virus involved in capsid assembly, J. Gen. Virol. 90 (2009) 392–397.

[66] L. Sompayrac, How Pathogenic Viruses Work, Jones and Bartlett Publishers, Sudbury, Massachusetts, 2002.

[67] G.K. Goh, A.K. Dunker, J.A. Foster, V.N. Uversky, Zika and flavivirus shell disorder: Virulence and fetal morbidity, Biomolecules 9 (11) (2019) E710, https://doi.org/10.3390/biom9110710.