

Feasibility of the vaccine development for SARS-CoV-2 and other viruses using the shell disorder analysis

Gerard Kian-Meng Goh,^{1,*} A. Keith Dunker,² James A. Foster,³ and Vladimir N. Uversky⁴

¹*Goh's BioComputing, Singapore 548957, Republic of Singapore (gohsbiocomputing@yahoo.com)*

²*Center for Computational Biology and Bioinformatics, Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 410 W. 10th St, HS5000, Indianapolis, IN46202, USA*

³*Department of Biological Sciences University of Idaho, Moscow, ID 83843, USA*

⁴*Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL, USA*

Several related viral shell disorder (disorder of shell proteins of viruses) models were built using a disorder predictor via AI. The parent model detected the presence of high levels of disorder at the outer shell in viruses, for which vaccines are not available. Another model found correlations between inner shell disorder and viral virulence. A third model was able to positively correlate the levels of respiratory transmission of coronaviruses (CoVs). These models are linked together by the fact that they have uncovered two novel immune evading strategies employed by the various viruses. The first involve the use of highly disordered “shape-shifting” outer shell to prevent antibodies from binding tightly to the virus thus leading to vaccine failure. The second usually involves a more disordered inner shell that provides for more efficient binding in the rapid replication of viral particles before any host immune response. This “Trojan horse” immune evasion often backfires on the virus, when the viral load becomes too great at a vital organ, which leads to death of the host. Just as such virulence entails the viral load to exceed at a vital organ, a minimal viral load in the saliva/mucus is necessary for respiratory transmission to be feasible. As for the SARS-CoV-2, no high levels of disorder can be detected at the outer shell membrane (M) protein, but some evidence of correlation between virulence and inner shell (nucleocapsid, N) disorder has been observed. This suggests that not only the development of vaccine for SARS-CoV-2, unlike HIV, HSV and HCV, is feasible but its attenuated vaccine strain can either be found in nature or generated by genetically modifying N.

Keywords: SARS; COVID; disorder; Coronavirus; HIV; vaccine; virulence; viral shell.

1. Introduction

1.1. SARS-COV-2 Vaccine

Since its outbreak in December 2019, a dangerous coronavirus (CoV), severe acute respiratory syndrome CoV-2 (SARS-CoV2), causing Coronavirus Disease (COVID-19) has spread rampantly with the dire consequences including large numbers of deaths and morbidities [1]. The SARS-CoV-2 spread has been so severe that many believe that it could only be kept in check with the discovery and availability of effective vaccines. While the successes in the discovery of vaccines for a large variety of viruses, including classical viruses, such as smallpox and rabies viruses,

© 2020 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.

provide grounds for greater hope, optimism and inspiration toward the discovery of COVID-19 vaccines, there are also nightmare scenarios, in cases, such as HIV (human immunodeficiency virus), HCV (hepatitis C virus), and HSV (herpes simplex virus), for which no vaccine has been found despite searches that span close to 40, 30, and 100 years, respectively. The polio vaccine development itself took 30-40 years, but those years were before the era of powerful modern molecular technology. It would therefore be unfair to make such comparison [3,5]. A question is then: Will the search for a SARS-CoV-2 vaccine be a nightmare as seen in HIV, or will it be a spectacular success, as in the case of rabies and smallpox? We shall see that the shell disorder analysis has much to say in this regard.

1.2. Shell disorder analysis of HIV and other viruses

In 2008, we reported that the use of artificial intelligence (AI) found some strange features in the outer shell (matrix) of the HIV-1, which was found to be very disordered [6]. We were not able to detect this feature in any other virus, despite a search among of a somewhat wide variety of unrelated viruses, such as influenza virus, rabies virus and the HIV's cousin, EIAV (equine infectious anemia virus) [6]. In subsequent years, similar levels of disorder can be found in very few other viruses, including HSV and HCV [2]. Both viruses are associated with sexual transmission, and no effective vaccine has been found for both. These cannot be explained by current the standard textbook paradigm [2].

In these and similar studies, the levels of protein intrinsic disorder [7-12] were measured for proteins constituting shell of each analyzed virus using a neural network-based predictor PONDR® VLXT [13,14]. This algorithm predicts the intrinsic disorder predisposition of each residue in a protein. A convenient yardstick to measure the level of disorder in a protein is PID (percentage of disorder). In the case of HIV-1, the matrix PID reaches the high level of 70% [2].

1.3. Spinoff projects including coronaviruses: Shell disorder and modes of transmission

Following the success of the HIV shell project, several spinoff projects based on the similar ideology were initiated. One of these spinoffs was the coronavirus project. Before the MERS-CoV outbreak in 2012, a shell disorder model was built to predict the mode of transmission of this virus based mostly on the levels of intrinsic disorder in its inner shell (N, Nucleocapsid) but also partly taking into account the disorder status of the outer shell (membrane, M) [15]. When the PIDs of M and N were measured for the variety of CoVs, the viruses were clustered into three groups mainly based on their N PID values. Those with the highest PIDs are those with higher respiratory but lower fecal-oral transmission potentials. Those with intermediate levels of N PIDs are the CoVs predicted to have intermediate levels of both respiratory and fecal-oral transmission potentials. The model was developed using knowledge of the behaviors of animal CoVs, particularly porcine CoVs from the veterinary community [3] and was later further validated using multivariate analysis [3,15,16].

In this model, SARS-CoV was placed in group B that contains CoVs with intermediate - respiratory and fecal-oral transmission potential and the results of the model were published in 2012 [15]. Upon the MERS-CoV (Middle Eastern respiratory syndrome CoV) outbreak, the characterization and classification of MERS-CoV had to wait until the time when the genome or proteome sequences of this virus became available. However, as soon as this information was released, it was used to analyze MERS-CoV. This analysis revealed that MERS-CoV belongs to the group C that contains CoVs with higher fecal-oral but lower respiratory transmission potentials [16]. The results further validate the reliability of the model as clinical data of MERS-CoV do show that it is not easily transmissible among humans via respiratory routes and is associated with camels, which are often farmed and thus allow for greater fecal-oral transmission [17].

Yet another opportunity to validate the shell disorder model came with the COVID-19 outbreak. Again, the model was consistent with existing data and placed SARS-CoV-2 in the same category as SARS-CoV; i.e., CoVs with the intermediate levels of both respiratory and fecal-oral transmission potentials [18]. There was, however, something noticeably odd about this virus. Analysis showed that with the exception for HCoV-HKU1, SARS-CoV-2 had the hardest outer shell (lowest M_{PID}) in our fairly wide variety of CoVs analyzed then. This means that SARS-CoV-2 is likely to resist the anti-microbial enzymes found in the saliva and mucus of the host and is also likely to survive longer in non-physiological environments [19,20]. Further search has found that such hardness is associated with burrowing animals, such as rabbits and pangolins that are in contact with buried fecal materials [21,22]. Furthermore, clinical studies have shown that COVID-19 patients shed large amounts of SARS-CoV-2 viral particles, which are far exceeding levels of viral particles shed by infected with SARS-CoV [23]. The hardness of the outer shell and the ability of the virus to resist the anti-microbial enzymes in body fluids could account for these observations

1.4. Yet another spinoff: Correlations between the inner shell disorder and virulence

Yet another spinoff from the parent HIV vaccine mystery project is the discovery of a correlations between the inner shell disorder and virulence in fairly diverse set of related and unrelated viruses including Nipah virus (NiV), flaviviruses, Dengue virus (DENV), and Ebola virus (EBOV) [24-28]. In this paper, we will address the feasibility of SARS-CoV-2 vaccine development based on the shell disorder models and discuss the evolutionary aspects of SARS-CoV-2 and other viruses.

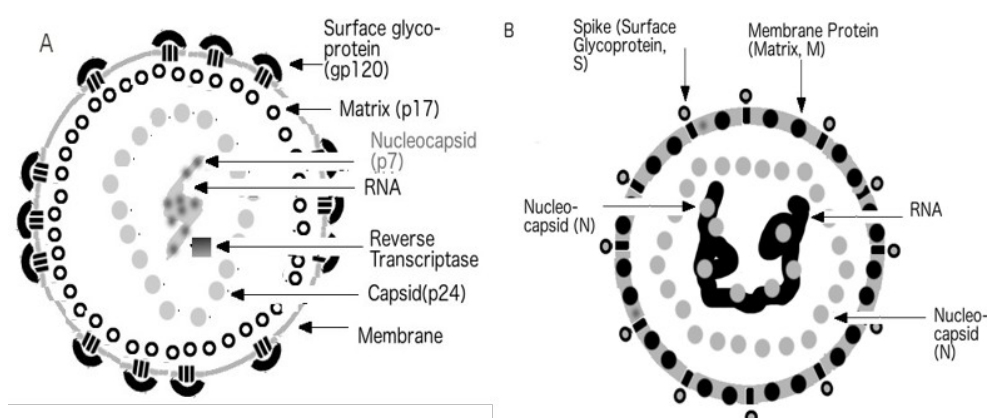


Figure 1. Virion Physiology A) HIV B) Coronavirus (CoV). (Figures reproduced with the permission of Gerard K. M. Goh 2017)

2. Results

2.1. Clustering of CoV based mainly on N_{PID}

As already mentioned, the CoV shell disordered model clustered CoVs into three statistically identifiable groups (ANOVA $p < 0.01$, **Table 1**), which correlated positively with the levels of respiratory transmission but negatively with the levels of fecal-oral transmission potentials. While the main contributing independent variable is the N_{PID} ($r^2=0.77$, $p < 0.01$), a slight increase in the correlation coefficient can be seen when both M_{PID} and N_{PID} are used as independent variables ($r^2=0.80$, $p < 0.01$). This implies that M_{PID} does contribute to the model even if slightly. **Figure 1** provides schematic virion physiology, with HIV and CoV as examples[3,4]. The inner and outer shells of CoV is the N and M proteins, respectively, as seen in **Figure 1B**.

Table 1. Categorization of coronaviruses by mainly N PID to predict levels of respiratory and fecal-oral transmission potentials ($p < 0.001$, $r^2 = 0.8$).

| Shell Disorder Group | Coronavirus | Accession Code (M Proteins) ^a | Accession Code (N Proteins) ^a | M _{PID} | N _{PID} | Remarks |
|----------------------|---------------------------|--|--|------------------|------------------|--|
| A | HCoV-229E | P15422(U) | P15130(U) | 23 | 56 | Higher levels of respiratory transmission lower levels of fecal-oral transmission |
| | IBV(Avian) | P69606(U) | Q8JMI6(U) | 10 | 56 | |
| B | Bovine | P69704(U) | Q8V432(U) | 7.8 | 53.1 | Intermediate levels of respiratory and fecal-oral transmission |
| | Rabbit | H9AA37(U) | H9AA59(U) | 5.7 | 52.2 | |
| | PEDV (Porcine) | P59771(U) | Q07499(U) | 8 | 51 | |
| | Canine (Resp.) | A3E2F6(U) | A3E2F7(U) | 7 | 50.5 | |
| | HCoV-OC43 | Q4VID2(U) | P33469(U) | 7 | 51 | |
| | SARS-CoV | P59596(U) | P59595(U) | 8.6 | 50.2 | |
| | HCoV-NL63 | Q6Q1R9(U) | Q6Q1R8(U) | 11 | 49 | |
| | SARS-Cov-2 | P0DTC5(U) | P0DTC9(U) | 5.9 | 48.2 | |
| | Bats ^b | A3EXD6(U) | Q3LZX4(U) | 11.2±5.3 | 47.7±0.9 | |
| C | MHV(Murine) | Q9JEB4(U) | P03416(U) | 8 | 46.8 | Lower levels of respiratory transmission higher levels of fecal-oral transmission |
| | Pangolin-CoV ^c | QIA428617(G) | QIA48630(G) | 5.6±0.9 | 46.6±1.6 | |
| | MERS-CoV | K0BU37(U) | K0BVN3(U) | 9.1 | 44.3 | |
| | TGEV(Porcine) | P09175(U) | P04134(U) | 14 | 43 | |
| | Canine (Ent.) | B8RIR2(U) | Q04700(U) | 8 | 40 | |
| | HCoV-HKU1 | Q14EA7(U) | Q0ZME3(U) | 4.5 | 37.4 | |

^aUniProt(U): <https://www.uniprot.org>; Genbank-NCBI(G): <https://www.ncbi.nlm.nih.gov/protein>

^b3 out of 4 bat-CoVs are in group B. Note: Large standard deviation can be seen for N_{PID} as denoted by “±”

^c2 out of 3 pangolin-CoVs are in group C. One is almost identical to SARS-CoV-2 in N PID. All samples were found to have high sequence similarities to the corresponding proteins found in SARS-CoV-2

2.2 Outer shell disorder is an indicator for the presence or absence of effective vaccines

While disorder at the inner shell is correlated with the mode of transmission, high outer shell disorder is associated with difficulties in finding effective vaccines. We shall see later that disorder at the inner shell (and sometimes at the outer shell as well) correlates with virulence.

Tables 2-3 summarize the disorder of the different shells of a variety of related and non-related viruses. There are no effective vaccines for HIV, HCV and HSV, which have abnormally high outer shell disorder. Conversely, viruses for which effective vaccines are available have ordered outer shells. These include rabies, yellow fever virus, smallpox virus and rotavirus [2-4,29]. The poliovirus has a capsid that is made up of a complex of several proteins, which are all relatively ordered. As aforementioned, the effective vaccines for polio have been available since the 1950s [4,5].

Table 2. Viruses and their descriptions. UniProt (<http://www.uniprot.org>) accession codes for shell proteins are given.

| Virus | Virus Type, Transmission | Outer Shell, Proteins (UniProt Accession)* | Intermediate Shell | Inner Shell |
|-------------------|--------------------------------------|--|----------------------|----------------------------|
| EIAV ^a | Retroviridae (RNA) | Matrix, p15 (P69732) | Capsid, p26 (P69732) | Nucleocapsid, p11(P69732) |
| FIV ^b | Retroviridae, Fights, Blood contacts | Matrix, p15 (P16087) | Capsid, p24 (P16087) | Nucleocapsid, p13 (P16087) |
| HIV-1 | Retroviridae Sexual | Matrix, p17 (P03348) | Capsid, p24 (P03348) | Nucleocapsid, p7(p17) |

| | | | | |
|--|---|--|---------------------------|--|
| HIV-2 | Retroviridae Sexual, Bite | Matrix, p17 (P04584) | Capsid, p24 (P04584) | Nucleocapsid, p7(P04584) |
| Variola/ Smallpox ^c | Poxviridae (DNA) Inhalation | Membrane, C9L(Q76U97), A14(P33839), F5(P33865) | | Core, VP8(Q0N570), 4A(Q0N532), 4B(Q0N539) |
| Rabies | Rhabdoviridae (RNA) Bites | Matrix, M(P25224) | | Nucleocapsid, N(P151979) |
| Poliovirus | Picornaviridae (RNA) Fecal-Oral | | Capsid, VP1-4 (P03302) | |
| Yellow Fever (YFV) | Flaviviridae (RNA) Insect | Membrane, M (P03314) | Capsid, C (P03314) | |
| Rotavirus | Reoviridae (RNA) Fecal-oral | Outer Capsid, VP7 (P21285) | Capsid, VP6 (P03530) | Capsid, VP2 (P12472) |
| SARS-CoV-2 | Coronavirus (RNA) Respiratory, Fecal-oral | Membrane (A3EXD6) | | Nucleocapsid (Q3LXZ4) |
| Hepatitis C (HCV) | Flaviviridae (RNA) Sexual | | | Core, p19, p21 (P26663) |
| Herpes Simplex Virus-2 (HSV-2) ^c | | Tegument, VP22-UL49 (A74K33), VP1/2-UL36 (I1UYK0), VP13/14- UL47 (A7LK25), VP16-UL48 (P68335), US3 (P13287) | Capsid, VP5 (p89442) | |

^aEquine Infectious Anemia Virus (EIAV)

^bFeline Immunodeficiency Virus (FIV)

^cOnly major shell proteins are considered

2.3. A disordered outer shell provides an immune evasion tactic: Viral shapeshifting

Evidently, as seen in **Tables 2-3**, some viruses like HIV evade the immune system via their disordered outer shell. The question is then: How do viruses do it? **Figure 2** summarizes the mechanism of immune evasion as seen in the case of HIV and its disordered matrix. The disordered matrix allows for motions that increase the movements of the surface glycoproteins such that the antibodies are not able to bind firmly to the virus. In the case of HIV, HIV antibodies can easily be found but neutralizing antibodies are difficult to find [2,3]. There are obviously various degrees of vaccine failures that depends on the level of outer shell disorder as we shall see in the case of FIV.

Table 3. Disorder levels (PID) of shell proteins. PIDs are arranged according proteins as stated in Table 2. Effective vaccines have been discovered for EIAV, rabies, polio, smallpox and rotavirus.

| Virus | PID of Outer Shell | PID of Intermediate Shell | PID of Inner Shell |
|-------------------------|-----------------------|-------------------------------------|------------------------|
| EIAV | 13±0.1 | 29±0.1 | 26±0.1 |
| FIV | 53.3±2 | 38.8±2.7 | 64.5±11.8 |
| HIV-1, SIVcpz | 56.5±10.8 | 44.5±2.6 | 39.5±3.0 |
| HIV-2, SIVmac | 51.5±2.5 | 26.6±2.9 | 46.5±0.1 |
| Smallpox ⁺ | 13±0.1 8±0.1 | | 19±0.1 4±0.1 12±0.1 |
| Rabies | 25.8±1.4 | 21.5±0.8 | |
| Poliovirus ⁺ | | 34±3.8 15.12±6.1 31.3±3.6 27±0.1 | |

| | | | |
|--------------------|----------------|----------------|-------------------------------|
| Yellow Fever (YFV) | 35.2 \pm 0.9 | 74.3 \pm 0.9 | |
| Rotavirus | 12.9 \pm 1 | 9.8 \pm 1.4 | 19 \pm 1.7 |
| SARS-CoV-2 | 5.9 \pm 0.1 | | 48.2 \pm 0.9 |
| HCV ⁺ | | | 52.5 \pm 0.5 48.5 \pm 0.5 |
| HSV-2 | 61 \pm 2 | | 18 \pm 0.1 |
| | 50 \pm 0.1 | | |
| | 38 \pm 1 | | |
| | 39 \pm 0.6 | | |
| | 37 \pm 0.1 | | |

*The standard error is denoted by “ \pm ”

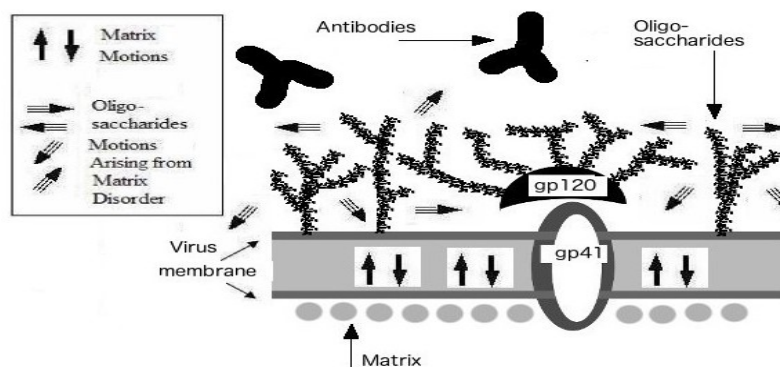


Figure 2. Viral “shape-shifting” immune evasion. Highly disordered matrix allows for greater motions at the glycoprotein that will prevent the antibodies from binding tightly to the virus. (Figure reproduced with the permission of Gerard K. M. Goh 2017)

2.4. SARS-CoV-2: Exceptionally hard shell (low M_{PID}) associated with burrowing animals and buried feces

While **Table 2** shows that it is difficult to find vaccines for viruses with extremely high levels of disorder in outer shell, **Figure 3A** reiterates that even though virtually all CoVs have relatively hard outer shell (for them, the fecal-oral transmission is generally an important route) SARS-CoV-2 has an exceptionally hard outer shell ie low M_{PID} . As a matter of fact, it has one of the hardest outer shells (among all the CoVs). A later search for CoVs with harder (low disorder) M protein came up with rabbit-CoV and pangolin-CoVs (see **Table 1**) [21,22]. There were obvious associations with burrowing animals that are likely in contact with buried feces. It should also be noted that while pangolin-CoVs were found to be closely related to SARS-CoV-2, the rabbit-CoV is not. Independent but parallel evolutions with similar evolutionary pressures are implied in the case of SARS-CoV-2 pangolin-CoVs, and rabbit-CoV [21,22].

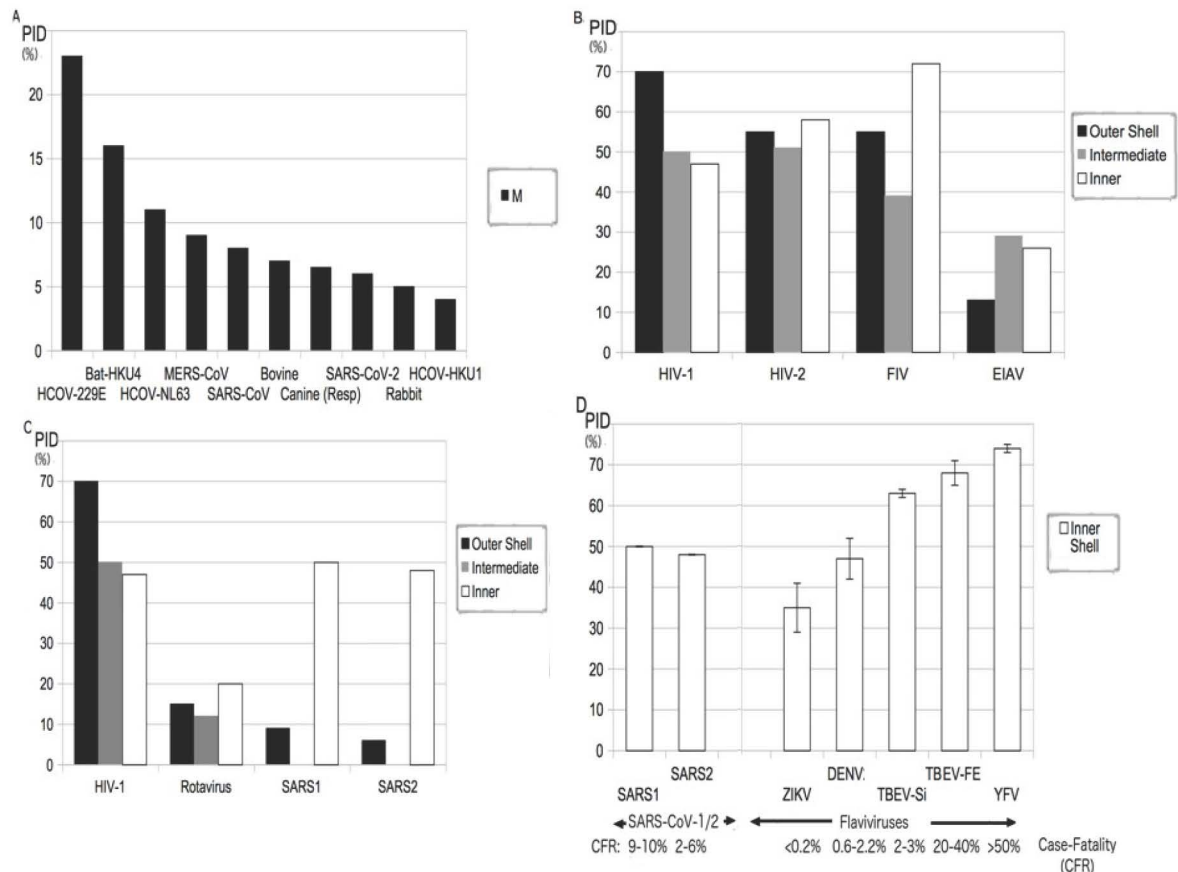


Figure 3. Quantification of shell disorder in different viruses. A) PIDs of M by CoV. SARS-CoV-2 has among the hardest M ie lowest M_{PID} . B) Shell disorder of selected retroviruses. C) Comparative shell disorder. SARS-CoV-2 (SARS2), 2003 SARS-CoV (SAR1). D) Inner shell disorder vs. virulence. Flavivirus capsid, C (inner shell. Case-fatality rate (CFR)

2.5. Behavior of the animal hosts matters in the evolutions of the viruses: EIAV vs. HIV

While it has been seen that higher outer shell disorder and the absence of effective vaccines are not just seen in retroviruses but also in other viruses, such as HCV and HSV as shown in **Table 2**, **Figure 3B** illustrates that not all retroviruses have high matrix disorder or lack effective vaccines, EIAV has ordered shells at all levels, and an effective vaccine for this virus was discovered in 1973 or before [30]. It should also be noted that HIV is predominantly sexually transmitted, whereas EIAV is transmitted by blood sucking horseflies, which hold its blood meal in their mouthpiece where the virus is exposed to the insect's saliva [2]. Unlike HIV, the EIAV needs the hard (low disorder) shells to protect itself against destructive effects arising from the anti-microbial enzymes found in the saliva. The “viral shape-shifting” immune evasive characteristic found in HIV is therefore absent in viruses, such as EIAV and rabies viruses. Experimental observations of outer shell disorder and the resulting immune evasion have been made [31-34].

2.6. Feasibility of developing attenuated vaccine strains for SARS-CoV-2

We have seen that SARS-CoV-2 looks nothing like the viruses, for which the effective vaccines are unavailable. Furthermore, it has one of the hardest outer shells among CoVs (**Figure 3A**). **Figure 3D** suggests that the attenuated vaccine strains can be obtained by lowering the disorder levels in its inner shell; i.e., N_{PID} . Flaviviruses are used here only as an example of the evidence of correlation [24,26]. While only correlations between flavivirus inner shell disorder and virulence

are shown, a variety of other viruses have been found to have such characteristics. These include NiV, EBOV and DENVs [24-28]. SARS-CoV-2 and the 2003 SARS-CoV do also provide hints of such correlation by having respective N_{PID} of 48% and 50%, while also having CFR of 2-6% and 9-10%, as seen in **Figure 3C**. The HIV and other viral shape-shifters exhibit a unique immune evading tactic as seen in **Figure 2**, SARS-CoV1/2, NiV, EBOV, and flavivirus use a different tactic, “Trojan Horse”, where the virus would replicate rapidly upon infection before the host immune system even recognizes its presence [3].

3. Discussion

3.1. Links between respiratory transmission, N (Inner shell) disorder, and virulence: Viral load in body fluids vs. vital organs

A puzzle arises when we inspect the data shown in **Table 1** and **Figure 3D**. In fact, **Table 1** tells us that there is a strong positive correlation between the inner shell disorder and respiratory transmission potentials of CoV, whereas **Figure 3D** reports positive correlations between inner shell disorder and virulence. What is then the connection between these two types of correlation? They are connected, because they are related to viral loads at different parts of the body. In order for respiratory transmission to be feasible, a minimal level of viral load in the saliva or mucus has to be attained. Similarly, death often occurs when the viral load in a specific vital organ exceeds a minimal threshold. What could then account for the discrepancy between viral loads in body fluids and vital organs? One answer is related to the ability of the virus to resist the anti-microbial enzymes found in saliva and mucus. Given this and the anti-microbial resistant hard (low disorder) outer shell of SARS-CoV-2, the significance of the observation that SARS-CoV-2 sheds high amount of viral particles without increase in virulence, when compared with SARS-CoV is reiterated.

3.2. Greater disorder in the inner shell proteins provide means for the more efficient replication of viral particles

It has to be kept in mind that inner shell proteins have varied but similar functions across virus species [4]. This accounts for the observation of correlations between inner shell disorder and virulence across virus species. For instance, N proteins of CoVs are involved in assembly of various viral proteins near or at the endoplasmic reticulum (ER) and Golgi apparatus in preparation for packaging [35,36]. Similarly, the C protein precursors of flaviviruses move towards the ER and bind to its membrane, where interactions of other viral proteins take place in the assembling of viral particles [4]. In the case of EBOV, the NP (nucleoprotein) helps forming a tube-like structure that assists in the transportation of viral proteins to the ER [37]. As for the NiV, the N protein binds to both L and P proteins to form the RNA polymerase [38], which is responsible for the viral RNA replication. The inner shell proteins therefore play important roles in the rapid replication of the viral particles. As we can see, instances of protein-protein/DNA/RNA interactions taking place are aplenty. The greater the disorder, the more efficient the inner shell protein is able to play its role in the replication of the virus as disorder provides for more effective protein-protein/DNA/RNA binding [7,18,39].

3.3 Two modes of immune evasion: “Trojan Horse” (inner shell disorder) and “viral shape-shifting” (outer shell disorder)

We have just described “Trojan Horse” immune evading strategy, where the virus replicates rapidly via inner shell disorder, before the host immune system could even recognize it. Oftentimes, such strategy backfires on the virus especially when the viral load overwhelms vital organs and thus killing the host. We have also described the other immune evading strategy,

“viral shape-shifting” in HIV as manifested in the outer shell disorder (**Figure 2**), there is evidence that HIV actually employs both strategies. This is complicated by the fact that the matrix (outer shell) assumes many of the roles that inner shell proteins of other viruses would normally have. These roles include embedding the proteins into the host membrane and assembling the viral proteins [4]. Evidence of HIV's adoption of such strategy can be seen by the fact over 90% patients infected by HIV-1 dies within two years of infection but the onset of symptoms (AIDS) for HIV-2 and FIV may take many years if at all [6,40,41]. Unsurprisingly, the maximal matrix PIDs of HIV-1, HIV-2 and FIV are 70%, 55% and 55% respectively (**Figure 3B**).

3.4. FIV, HIV-1 and HIV-2: Similarities and differences

As it was already mentioned, the “viral shape-shifting” immune evading strategy requires high outer shell disorder, as seen in HIV, HSV and HCV. This, too, presents an enigma, as HIV-1, HIV-2, and FIV have different degrees of outer shell disorder (70%, 55%, 55%, see **Figure 3B**). HIV-1 is spread globally via sexual transmission. While HIV-2 is also predominantly sexually transmitted, it is mainly found in parts of Africa near its rainforest reservoir of the monkey, sooty mungabey, which replenishes the virus by bites and other human interactions [41]. Similarly, FIV is predominantly spread through fights and subsequent blood contacts [40]. These could explain the discrepancies in levels of matrix disorder between HIV-1 and FIV/HIV-2.

3.5. FIV vaccine enigma: Questionable efficacy

While there are apparent differences in evolution and matrix disorder of EIAV, HIV-1, and FIV, there are also differences in the successes with respect to the search for their vaccines. The search for an HIV vaccine has been ongoing for nearly 40 years with abysmal failures, but an effective vaccine for its horse cousin, EIAV, was discovered in 1973 or before [30]. It was tested on 60 million horses and was able to deflect an oncoming pandemic. Needless to say, the matrix PID of EIAV is 13%, compared to the HIV-1 M_{PID} of 70%. A more enigmatic story can be found in the case of FIV. A ray of hope came in 2002, when a FIV vaccine became commercially available. It initially boasted of 82% efficacy against the FIV subtype A, but was later shown to be totally ineffective against strains from countries, such as United Kingdom. It was shown to provide only 58% protection for cats in Australia [42]. Finally, the vaccine was later withdrawn from the market in USA and Canada partly because of its questionable efficacy [40]. This is consistent with our data showing that FIV (matrix PID = 55%), like HIV-2, has more moderate matrix disorder than HIV-1.

4. Conclusions

4.1. Development of the SARS-CoV-2 vaccine is feasible and vaccine strains can be found in nature

The nightmare scenario in the frantic search for the SARS-CoV-2 vaccine would be that it all ends up like the search for the HIV vaccine or, even worse, FIV vaccine. However, we can all have a sigh of relief as the shell disorder models are unable to detect any similarities between SARS-CoV-2, HIV-1 and FIV in terms of the outer shell disorder or peculiarities of evolution with respect to their modes of transmission. The outer shell disorder of SARS-CoV-2 and SARS-CoV resembles more viruses, like rotavirus, for which there are effective vaccines. Unlike rotavirus that is solely reliant on fecal-oral routes, SARS-CoV-2 has a somewhat disordered inner shell. The presence of such feature is necessary for most viruses with respiratory transmission potentials, because, as already explained, a minimal viral load in the mucus or saliva is required for transmission. The higher inner shell disorder also provides means for the “Trojan horse” immune evasion. Because of this, strategies involving attenuation of the SARS-CoV-2 by creating N with greater order levels can be contrived. In fact, a previous study has suggested that a SARS-CoV-2 precursor could have entered the human population

via pangolins in 2017 or before as an attenuated mild virus as a result of the peculiarities of the behaviors of pangolins [21,22]. Therefore, the disorder analysis not only suggests that vaccine development for SARS-CoV-2 is viable but also points out that the attenuated vaccine strains can already exist in nature.

5. Materials and Methods

As already mentioned, protein intrinsic disorder is an important concept that can be used for various analyses of proteins. It basically refers to the protein regions or entire proteins that have no unique 3D structures. Disorder in proteins plays an array of significant roles, such as the recognition of binding sites [7-12]. AI has been successfully employed to recognize disordered regions. For instance, PONDR® VLXT (www.pondr.com) [11,12,43] deploys neural networks to recognize such regions, as it was trained using known disordered and ordered sequences. PONDR® VLXT has been successfully used in the study of structural proteins of a variety of viruses including HIV, HSV, HCV, NiV, EBOV, 1918 H1N1 influenza A virus, CoVs, DENV, and several flaviviruses, e.g., Yellow fever virus (YFV), and Zika virus (ZIKV) [1-3,6,15,16,21,22,24-28,44]]. The reason that PONDR® VLXT is highly suitable for the studies of structural proteins of viruses has to do with its sensitivity in detecting local sites for potential protein-protein/DNA/RNA/lipid interactions [45]. A useful ratio used in this study is PID (Percentage of Intrinsic Disorder), which is defined as the number of residues predicted to be disordered divided by the total number of residues in the protein and multiplied by 100%. The value of this parameter provides an estimate of the extent of disorder in the protein of interest. A relational database was built using MySQL, JDBC and JAVA. A JAVA program imports the sequence and disorder information into the database [44]. Sequence information are obtained from UniProt (<http://www.uniprot.org>) Multivariate analyses were done using R-statistical package [46].

References

1. G. K. Goh, et al. (2020) *Biomolecules*. 10: 331.
2. G. K. Goh, et al. (2019) *Biomolecules*. 9: 178.
3. G. K. Goh, *Viral shapeshifters: Strange behaviors of HIV and other viruses* (Simplicity, 2017).
4. N. H. Acheson, *Fundamentals of molecular virology* (Wiley, 2007).
5. D. M. Oshinsky, *Polio: An American story* (Oxford University Press, 2005).
6. G. K. Goh, et al. (2008) *Vir J*. 5: 126.
7. P.E. Wright, et al. (1999) *J Mol Bio*. 9293: 321-331.
8. V.N. Uversky, et al. (2000) *Proteins* 41: 415-427.
9. A.K. Dunker, et al. (2000) *J Mol Graph Model* 19: 26-59.
10. P. Tompa. (2002) *Trend Biochem Sci*. 27: 527-533.
11. A. Hatos, et al (2020) *Nucleic Acids Res*. 48(D):D269-D276.
12. A.K. Dunker et al (2015) *Semin Cell Dev Bio*. 37:44-55.
13. E. Garner, et al. (1999) *Genome Inf*. 10: 41-50.
14. P. Romero, et al. (2001) *Proteins*. 42 38-48.
15. G. K. Goh, et al. (2012) *J Pathog*. 2012: 738590.
16. G. K. Goh, et al. (2013) *PloS Curr*. 5.
17. M. Ferguson, et al. (2014) *Lancet Infect Dis*. 14: 93-94.
18. G. K. Goh, et al. (2020) *Microb Pathog*. 144: 104177.
19. A. M. Cole, et al. (1999) *Inf immun*. 67: 3267-3275.
20. D. Malamud, et al. (2011) *Adv Dent Res*. 23: 34-37.
21. G. K. Goh, et al. (2020) *Preprints*. 2020060327.
22. G.K. Goh et al (2020) *J. Proteome Research*. 2020.
23. R. Wolfel, et al. (2020) *Nature*. 581:465-469.

24. G. K. Goh, et al. (2016) *Mol Biosyst.* 12: 1881–1891.
25. G. K. Goh, et al. (2015) *Mol Biosyst.* 11: 2227–2344.
26. G. K. Goh, et al. (2019) *Biomolecules.* 9: e710.
27. G. K. Goh, et al. (2020) *Microb Pathog.* 141: 103976.
28. G. K. Goh, et al. (2020) *Preprints.* 2020050116.
29. J. Angel, et al. (2007) *Nat Rev Microbiol.* 5: 529-39.
30. H. Wang, et al. (2017) *Oncotarget.* 9: 1356-64.
31. N. Kol et al (2007) *Biophys J.* 92:1777-1783.
32. H. Pang et al (2013) *Retrovirology.* 10:4.
33. Y. Ohori et al (2014) *Biochim Biophys Acta.* 1844:520-526.
34. F. Caccuri et al (2014) *J Vir.* 88:5706-5717.
35. S. Lu et al (2020) *BioRxiv.* 2020.07.31.228023.
36. T. McBride, et al. (2014) *Viruses.* 6: 2991-2018.
37. S. Wantanabe, et al. (2006) *J. Vir.* 80: 3743-51.
38. J. Habchi, et al. (2012) *Mol Biosyst.* 8: 69-81.
39. M. Macossay-Castillo, et al. (2019) *J Mol Bio.* 431: 1650-70.
40. M. Scherk, et al. (2012) *J Fel Med Surg.* 15: 785–808..
41. J. Goudsmit, *Viral sex: Nature of AIDS* (Oxford University Press, 1997).
42. B. Sahay, et al. (2018) *Viruses.* 10: 277.
43. X. Li, et al. (1999) *Genome Inform Ser Workshop Genome Inform.* 10: 30–40.
44. G. K. Goh, et al. (2008) *BMC Genomics.* 9 Suppl 2: S4.
45. C.J. Oldsfield et al (2005) *Biochemistry.* 44:12454-12470.
46. R Core Team. (2019) *A language and environment for statistical computing.*