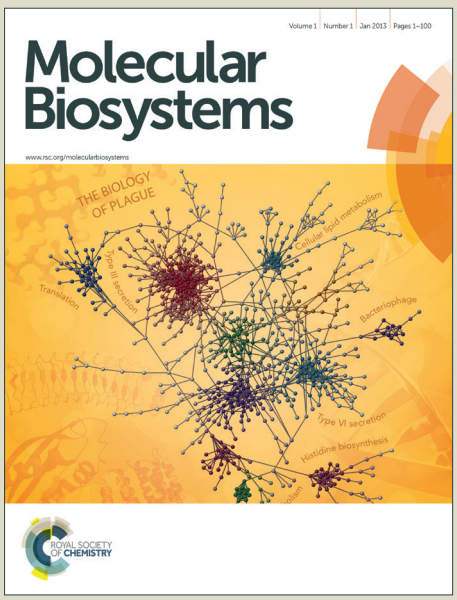
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Shell Disorder, Immune Evasion and Transmission Behaviors among Human and Animal Retroviruses

Gerard Kian-Meng Goh[1*], A. Keith Dunker[2], Vladimir N. Uversky[3]

[1]Goh's BioComputing, Singapore, Republic of Singapore

[2]Center for Computational Biology, Indiana and Bioinformatics, University School of Medicine, Indianapolis, Indiana, USA

[3]Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, Florida, USA

[3]Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow region, Russia

[*]Corresponding author

Email addresses:

GKMG: gohsBioComputing@yahoo.com

AKD:: kedunker@iupui.edu

VNU: vuversky@health.usf.edu

Abstract

This study involves measurements of percentages of intrinsic disorder (PIDs) in the GAG protein shells of various retroviruses. Unique patterns of shell protein disorder can be seen especially when GAG proteins (matrix M, capsid C, and nucleocapsid N) of primate and non-primate retroviruses are compared. HIV-1 presents the most unique pattern of disorder distribution with generally high levels of disorder in all three proteins, while EIAV (PIDs:: 26, 29, 13) is diametrically different from HIV-1 (N C M PIDs: 39.5±3.0, 44.5±2.6, 56.5±10.8). The HTLV viruses (CPID: 32.8±3.4 resemble HIV-2 (C PID:26.6±2.9) with a moderately disordered capsid. Totally distinct patterns, however, are seen for the non-primate retroviruses. They generally have highly disordered nucleocapsids (PID > 65%) and more ordered outer shells especially matrix. These characteristics might be attributed to the differences in the way the retroviruses are transmitted, with non-primate viruses having greater non-sexual transmission components such as oral-fecal transmission. These differences are also evolutionarily related to the ways the viruses evade the host immune systems, and thus, have implications for oncolytic virotherapy and animal models in vaccine research. The importance of protein shell disorder in immune evasion, as related to the case of HIV-1, and the difficult search for its vaccines are highlighted.

.

.

Introduction

In a previous paper[1], the protein intrinsic disorder prediction techniques were used to evaluate the differences in the matrix (outer shell) proteins of HIV-1, HIV-2, and EIAV. The use of a simple measure, percentage of intrinsic disorder (PID) revealed the highest levels of disorder in the HIV-1 matrix, compared to those in HIV-2 and EIAV, with EIAV having the lowest PID score. The trend was attributed to the differences in the modes of transmission of the related viruses, since it is known that the shells of viruses play a role in protecting the virion from damage by the environment [1-5]. New questions have, however, arisen. How are the other shell proteins (i.e. capsid and nucleocapsid) affected in terms of intrinsic disorder? There is also a wide range of other retroviruses, especially those with non-primate hosts that need to be looked at from this angle. In fact, primate and non-primate viruses tend to have differences in their modes of transmission and likely to have different hardness in their shells.[2, 3]

Another question that the paper will attempt to address is the role of viral shell disorder in immune evasion. One of the greatest medical puzzles for the last few decades has to do with the failure to find an effective HIV-1 vaccine despite tremendous amount of resources spent, even though an effective vaccine for its cousin, EIAV, has been available for a long time. This enigma will be addressed using our results

Yet another important note has to do with the fact the EIAV is not alone in its differences with HIV-1 especially with respect to disorder in their outer shells. Strangely, with a large variety of viruses inspected, very few viruses have disordered outer shells that are even comparable to that of HIV-1[1-3, 6]. An exception is the herpes simplex virus (HSV), which has some resemblance to HIV-1 in its disorder at the outer shell[7]. Enigmatically, the search for an effective vaccine has also been shrouded with failures to this date just like HIV but arguably to a somewhat lesser extent[8]. In this paper, we will extend the disorder analysis to HIV's relatives with respect to the GAG proteins to see if HIV-1 is any way unique in such respect too.

Table 1 represents some of the known retroviruses and their transmission modes[9] Table 1 shows that even within a given genus, there can be rather different transmission modes. Furthermore, while some retroviruses have similar modes of transmission, they are likely to be distinct by some subtle differences in the levels of the various transmission potentials.

While attempts to give greater emphasis on immune evasion are being made, given its greater medical implications, the research in this paper is by no mean the first attempt to link shell disorder to viral transmission patterns. In fact, the main tool used in this paper, PONDR®-VLXT[10, 11] has been applied to proteins of a large variety of viruses including some highly virulent ones such as HIV[1, 6], the SARS-CoV (Severe Acute Respiratory Syndrome Coronavirus)[2], MERS-CoV (Middle Eastern Respiratory Syndrome Coronavirus[3], FMDV (Foot and Mouth Disease Virus)[5] 1918 H1N1 and H5N1[12] viruses. A previous attempt to do so with coronaviruses has been met with surprising success. Even before the identification of the MERS-CoV, we were able to use the same methodology as described in this paper to categorize coronaviruses by groups according to their shell disorder, with the SARS-CoV belonging to a group of viruses with both moderated respiratory and oral-fecal transmission potentials[2]. Upon the availability of the genomic sequences of the MERS-CoV, the model was able to accurately predict that MERS-CoV had to fall into a separate group that is distinct from the one that contains the SAR-CoV[3]. This group that the MERS-CoV falls into consists of viruses with higher oral-fecal transmission potential but with relatively lower respiratory component, which is highly consistent with experimental and clinical data[13, 14].

Material and Methods

Protein intrinsic disorder

An important concept used in this paper is protein intrinsic disorder. It arises from the observation that while many proteins have rigid structures that can be easily determined, many

biologically active proteins, known as intrinsically disordered proteins, lack stable structures [15].. Such lack of stable structure is inherently linked to the sequences and functions of these proteins. In fact, the presence of hydrophilic residues tends to give rise to disordered regions, while hydrophobic residues generally define more ordered areas [2, 3]. There is a wide array of computational tools developed for the protein intrinsic disorder prediction. The main tool used in this study is the disorder predictor, PONDR®-VLXT (http://www.pondr.com), which is a neural network that is fed with the protein sequences and predicts which residues of the proteins are expected to be disordered [10, 11, 16]. The use of this predictor has been effective in providing insights to the roles of many crucial viral proteins[2, 3, 5-7, 12]. A useful standard for measuring levels of disorder in a protein is PID, which is defined as the number of residues predicted to be disordered (i.e., possessing the disorder score above 0.5) divided by the total number of residues in a given protein.

PONDR®-VLXT

PONDR®-VLXT was the earliest disorder predictor built[11, 17]. Since then, many predictors that are more :"accurate" as assessed by CASP (Critical Assesment of Structure Prediction)[18] have been developed. The determination of "accuracy" is based on the criteria that a protein or segment of a protein is considered disordered if its coordinates are not availability in NMR and X-ray crystal structures of higher resolutions[18, 19]. While such a criteria has led to newer and powerful predictors that are effective crystallographers' tools, in the identification of highly unstructured proteins that are imposible to crystallized, such a narrow definition of disorder has also been a subject of much controversy[19-23]. There are many good reasons for this, but only a few will be discussed here. Firstly, the process of crytalization could involve rhe exposue of proteins to harsh non-physilogical considtions such as high salt contents, so as to force the proteins to crystalize[24]. Then, there is also the issue of induced folding, This involves proteins

becoming ordered only upon contacts with their binding partners, which could include DNA/RNA and other proteins[20, 23, 25]. Such would come into conflict with the above-mentioned definition when, in realiy, there are many proteins that are in contact with their binding partners only occasionally or even rarely in nature[23, 26]. Furtherrmore, even among proteins that have been crystallized, flexible segments have also been detected[1, 22]. The narrow definition for disorder would force predictors to miscategorize many of the mentioned regions and proteins as ordered.

As mentioned, the definition of disorder used to evaluate predictors has encouraged the design of newer and more powerful predictors well suited as crystallographers' tools in the identification and analysis of extremely disordered proteins that are impossible to crystallize. Unfortunately, the proteins that are the focus of this paper, the GAG shell proteins, do not fall into this category i.e. extremely disordered proteins, and, not surprisingly, have been observed to be unsuitable for analysis using many of the newer predictors (unpublished data). In fact, all viral shell proteins should have some levels of order so as to provide a more ridgid encasement in order to protect their virions from environmental and physiological damages[2, 6]. Such shells do vary, however, in disorder depending on the way that the viruses had evolved. The sucesses of PONDR®-VLXT in the study of viral shell proteins from a variety of viruses, as mentioned above, are therefore not without reasons. Yet more evdence of the suitability of PONDR®-VLXT can be seen by the fact that VLXT has been found to be the most accurate for detercting protein-protein interactions[19-21].

Other tools and accessories

Relational databases were built to capture the information from PONDR®-VLXT, UNIIPROT (http://www.uniprot.org) and PDB (Protein Data Bank: (http://www.ncbi.nlm.nih.gov/structure/) [16]. Various JAVA programs were written to allow automated inputs, data entry and to generate codes readable by Jmol, which is a graphics program that is used to illustrate the three-

dimensional protein structures[27]. Representative accession codes used in our analyses are shown in Table 2. R statistical package was used to do ANOVA (Analysis of Variance). Also NCBI-BLAST (Basic Local Alignment Search Tool)[28] was used to determmine the presence of sequence alignment of RSV (Rous Sarcoma Virus) and HIV-1 nucleocapsid proteins.

Protein selections

As mentioned, the GAG proteins are the focus of this paper. The main GAG proteins are the matrix (outer shell), the capsid (intermediate) and the nucleocapsid (inner). These proteins have individual names as related to their virus types as seen with parentheses in Table 2. The GAG gene is found in all retrovirus[9, 29]. For this reason, a disorder study of the three proteins provides for an opportunity to do comparative analyses across the viral relatives. Because shell proteins are associated with protective and immune0evasive roles, comparative analysis of viral behaviors becomes an additional option available. Therefore, to make the comparative analysis more feasible, retroviral relatives that are more commonly known are chosen as seen in Tables 1-2. While it is obviously more difficult to find information involving animal retroviruses than that of human, the animal retroviruses listed in Tables 1-2 are those with information that is more easily available.

Result

Shell disorder (PID) proximities

As mentioned above, an important benchmark used to measure the level of disorder in a protein is its PID. Table 3 represents PID levels evaluated in shell proteins of various retroviruses. The viruses are grouped together by genus and, hence, genetic proximity. The grouping of PIDs by genus allows to us see that there is no necessary correlation between predicted disorder and genetic proximity. (i.e. genus). In fact, we will argue that the

discrepancies arise from the differences in transmission modes of the different viruses since a rigid encasement (more ordered shells) is required if the virus is likelier to be exposed to harsher environments. Sample size for each protein is given in Table 3. The study involves 10 virus types and approximately 34 virus strains with a total of 104 protein samples.

Protein shell disorder and transmission modes

While many genetically close viruses that have similar transmission modes and therefore possess similar patterns of disorder in their shell proteins, the disorder patterns at the shells is expected to be completely different if these viruses have different transmission modes , regardless of their genetic proximity. Such characteristics have been shown in the past and will be further revealed in this paper. This can be seen in the case of EIAV and its genus, lentivirus, which basically means that it is more closely related to HIV-1, FIV and HIV-2 (Table 1). As seen in Table 3, EIAV has a distinct pattern of disorder distribution in its shell proteins that is totally different from that of its cousins, especially the HIVs and FIV. It is likely that this difference in disorder distribution can be related to the differences in the transmission modes of these retroviruses. In fact, EIAV is transmitted by a blood–sucking horsefly, whereas the other lentiviruses (such as HIV) are spread by sexual contacts, intravenous drug use and breast feeding[30-32].

Table 3 shows that the nucleocapsid PIDs of non-primate viruses are extremely high whereas PIDs of their outer GAG proteins are relatively low, even when they are compared to the corresponding values calculated for primate retroviruses. An explanation for this behavior can be related to the fact that all the shell proteins have to play a role in protecting the virion from damage and yet have to possess sufficient flexibility (disorder) to conduct their other functions. Therefore, if a non-primate virus has harder outer shells as a result of the need to protect its virion from damage during oral-fecal spread, then the nucleocapsid might play a

compensatory role by being more flexible. The opposite possibility is also true, especially for human retroviruses with large contribution of the sexual transmission and a virtually non-existent oral-fecal transmission mode. Also, a large level of disorder at any of the shell proteins is likely indicative of the different modes of immune evasion.

HIV versus EIAV: Transmission and protein intrinsic disorder

Figure 1 gives a closer look at the peculiarities of disorder distribution in the shell proteins of lentiviruses. The disorder of EIAV matrix is different from that of its HIV cousins and the matrix disorder of HIV-1 is different from that of HIV-2 [1]. Table 3 and Figure 1 show that the patterns of disorder distribution in nucleocapsid, capsid and matrix are different for the various lentiviruses. Notably, EIAV is unique among this set of lentiviruses, by being transmitted by insect, it also has a unique pattern of low disorder for all its GAG proteins, unlike its cousins that are transmitted sexually. It should also be noted that "ALL" in Figure 1 refers to the means and standard errors of the corresponding proteins in viruses currently found in the database as described in a previous paper[6], which includes a wide variety of retroviral and non-retroviral animal viruses. This is to allow us to observe the high levels of disorder that can be found in some retroviruses such as HIV-1 and, conversely, the low levels of disorder in EIAV.

EIAV and HIV-1: Insect transmission and IV drug abuse

The EIAV is spread mainly by horseflies that carry their fresh virus-laden blood meals on their mouth piece before biting previously uninfected hosts[33]. Moreover, contamination of syringes in veterinary clinics has been known to cause EIAV spread. Given these, one could easily be tempted to assume that EIAV spread is similar to viral transmission by contaminated intravenous (IV) drug injections, which are also a mode of transmission in HIV-1. Such analogy would involve, however, being oblivious to the fact that EIAV is likely to be constantly exposed

to the insect's saliva when held as a blood meal near the mouth of the horsefly[34]. In a stark contrast, our data and model emphatically tell us that viruses such as EIAV, which are in constant contact with saliva, face totally different evolutionary challenges from those that are spread mainly by IV drug abuse. Our results, on the other hand, can be supported from a biochemical and physiological viewpoint since saliva contains digestive enzymes such proteases and glycosidases that are potentially harmful to viruses in general[3, 35]. Viruses that are constantly exposed to saliva, therefore, need to evolve in ways to protect their virions by having hard shells especially at the outer layers. Such is essentially consistent with our other data on other insect-related and saliva-associated viruses such rabies, yellow fever virus, dengue viruses and vesicular stomatitis virus(VSV) (unpublished data found in database mentioned in previous papers [3, 6]).

EIAV: Low disorder at all shell levels

A subtle but important feature that needs to be looked upon with attention is the fact that, as we can see, as in Figure 1 and Table 3, the shell disorder of EIAV is low at all levels, not just at the outer shell (Matrix PID: 13 $\pm$ 0.1; Capsid PID: 29$\pm$0.1; Nucleocapsid PID: 26$\pm$0.1 . This is in stark contrast not only to the other retroviruses sampled in this paper but also to most non-related viruses (unpublished data in database described in[3, 6]) that are transmitted by insects. Many of these arboviruses have low PID at the outer shell but not necessarily at the inner shells. It should also be noted that while the capsid and matrix proteins of HIV-1 and EIAV have already been discussed to a limited extent in a previous paper, their nucleocapsid PIDs were, however, not mentioned[6]. This topic has to be revisited in a later section of the paper as it has important implications for viral immune evasion and the interpretation of our disorder model.

HIV-1 versus HIV-2: Transmission modes and shell disorder

Another peculiarity can be seen among the HIV viruses. The difference in disorder between HIV-1 and HIV-2 is not just at the matrix as seen in the previous paper [1] but also extends to the capsid. The means of capsid (26.6±2.9) and matrix (51.5±8.5) PIDs of HIV-2 are generally lower than those of HIV-1 (Capsid PID: 44.5±2.6, Matrix PID: 56.5±10.3), even though a huge variance is found for HIV-1. An explanation for this difference can be found in the subtle differences in the way HIV-1 and HIV-2 spread. While sexual contact is the main mode of transmission for both viruses, HIV-2 is found only in limited regions around the world, namely in parts of Africa, where it is in close proximity to its reservoir, such as SIVsm in sooty mangabey monkeys [30, 36, 37]. For this reason, bush-meat consumption and bites are likely modes of cross-species transmission[38]. The disorder data seem to reaffirm this important evolutionary bottleneck seen in HIV-2, which has remained essentially in close proximities to its simian reservoir as mentioned.

FIV and HIV-2 have lower capsid disorder than that of HIV-1

We have mentioned above that non-primate viruses have patterns of disorder that are different from those of human and other primates. The FIV is seen as having a higher nucleocapsid PID (71.3±0.6) and, more importantly, a relatively moderate PID for its capsid (FIV: 37.2±0.5; HIV-1: 44.5±2.6:, see Figure 1). While sexual transmission is possible FIV is commonly transmitted by cat fights and bites, unlike HIV-1 [33, 39].. The fact that HIV-2 has lower capsid(HIV-2: 26.6±2.9) and matrix (51.5±8.5) PIDs when compared to those of FIV (Capsid: 37.2±0.5; Matrix: 54.3±1.2) , suggests that there are also differences in the common routes of transmission, which are exemplified by the regular consumption of monkey bush-meat by humans as in the case of HIV-2/SIVmac[30].

Bites and salivary exposures: HIV-1, FIV, EIAV and HIV-2

We have seen that FIV is commonly spread during cat fights that involve blood contacts and bites, which could account for its lower capsid disorder, It should, however be noted that there are cases when HIV-1 is spread between humans during fights that often involved bites, but this is not a common form of transmission, unlike FIV[40]. Given the patterns of the shell disorder especially for HIV-1, the model tells us that it is possible for HIV-1 to be found in saliva, but, depending on the strain, it is not likely to remain in the saliva for a longer period than what can be seen in FIV or HIV-2. It would, however, initially seem contradictory that HIV-2 is not usually spread by fights and bites, at least, among humans. As mentioned above, there is evidence that HIV-2, unlike HIV-1, faces an evolutionary bottleneck as seen by the fact that infections even among humans tend to be in close geographical proximity to its natural reservoir of SIVsm among the sooty mangabeys in West Africa. Cross specie transmissions between humans and non-human primates often involve bites and animal attacks[30, 38]. On a related note, it needs also to be reiterated that the low disorder in all three EIAV shells is likely a necessity in order to protect the virion from damage, given that a small quantity of the virus is held for a prolonged period as blood meal near the insect's mouth with constant exposure to the saliva containing harmful digestive enzymes

HTLV/STLV viruses are more similar to HIV-2/SIV in matrix disorder than FIV or EIAV: Behaviors of primates

The characteristically higher matrix PIDs can generally be seen not only in HIV, but also in other members of the primate retroviruses. This is illustrated by Figure 2 which compares shell proteins of the two sets of viruses from different genii (i.e., lentiviruses and deltaviruses,

see Table 1). Even though HIV (HIV-2 Matrix PID: 51.5±8.5) and HTLV (HTLV-2 Matrix PID: 46±0.1) belong to different genii, they have greater resemblance with respect to the patterns and ranges of the PIDs. There are two interrelated factors that need to be taken into account to explain the trend. Firstly, both HIV and HTLV have primate hosts. Secondly, HTLV and HIV viruses have similar transmission routes, such sexual contacts and breast feeding [4, 9, 30, 41]. There is therefore no evolutionary pressure on the shell proteins in order to make them more suitable to other transmission modes, such as the oral-fecal transmission mode that often present in non-primate viruses.

HTLVs are more similar to HIV-2 than HIV-1 in capsid disorder

There is however one observation that needs to be emphasized. The patterns and levels of disorder in the shell proteins of HTLVs are all much more similar to those of HIV-2 than to those of HIV-1 (Capsid: 44.5±2.6). This is especially the fact for the two outer shells, the capsid and matrix, where the capsid PIDs of the HTLV viruses (HTLV2: 22±0.1) reach low levels just like HIV-2 (26.6±2.9). An explanation for these characteristics can be attributed to the evolutionary similarity of the HIV-2/SIVmac and HTLV/STLV. Just as HIV-2 replenishes itself in the human population from sooty mangabey infected with SIV via bites and consumption of infected bush-meats, HTLVs is likely have similar relationships with their corresponding STLVs [42]. The data that are shown in Figure 2, seem to provide some support for the links. There is, in fact, genetic evidence pointing to regular cross-species transmissions in HTLV/STLV [30, 43-45] among non-human primates and between human and non-human primates.

Primate versus non-primate viruses in shell disorder: Transmission behaviors

Figure 3 illustrates the differences between primate and non-primate animal retroviruses and shows that the patterns of shell disorder of non-primate retroviruses are obviously different

from those of primate retroviruses. For example, for the non-primate viruses, there are distinctly higher levels of disorder for nucleocapsid (Oneway ANOVA, $p < 0.05$, $F= 76.7$) and lesser disorder for capsid and matrix proteins. The reason for this trend is likely to lie in the greater probability of non-primate viruses [46] to possess higher levels of oral-fecal transmission mode, which could be seen in the characteristic of having hard outer shells but a soft nucleocapsid.

While statistical analysis (Oneway ANOVA, $p < 0.05$, $F= 76.7$) provides evidence for differences in nucleocapsid disorder between primates and non primate retroviruses, one could also argue that genetic proximity of the viruses contributes to the findings. However, a closer look at the data will tell us otherwise. The Oneway ANOVA was conducted with FIV and HIV-1 as different virus classifications even as they are obviously both genetically close since they come from the same genus, lentivirus (Table 1). More intriguingly, a comparison of their nucleocapsid PIDs (Figure 1) reveals that large disorder differences can be found in not just between FIV and HIV-1 (PID: 71.3 ± 1.5 Vs 39.5 ± 3.0) but also between FIV and HIV-2 (PID: 71.3 ± 1.5 Vs 46.5 ± 9.25). Conversely, FIV, RSV and FeLV, which are non-primate retroviruses, are not closely related but yet have similarly high nucleocapsid PIDs (71.3 ± 1.5, 68.5 ± 10.5, 73.9 ± 5.1, see Figures 3-4, Table 1). Likewise, HTLVs and HIVs are deltaretroviruses and lentiviruses respectively but yet have relatively similar nucleocapsid PIDs, in contrast to that of the non-primate lentivirus, FIV (See Figure 3). This, we argue, is further evidence for the differences in the evolutionary pressures of viruses of primate and non-primate origins via transmission modes regardless of genetic proximity.

Another Oneway ANOVA was performed using the matrix (M) PID among primate and non-primate retroviruses. It yielded highly statistically significant results especially when EIAV and FIV were excluded ($p < 0.01$, $F = 48.5$). Our model is also detecting an unusually disordered outer shell (M) in FIV similar to those found in HIV-1, but not among other non-primate retroviruses. Apparently, sexual intercourse is likely be an evolutionarily important mode of

transmission for FIV even if it is not a commonly observed mode among cats[33, 39] as we shall see in the next paragraph..

FIV versus FeLV shell disorder: Non-casual vs. Casual contacts

Yet another clue to support this explanation can be found by looking at the differences in both PID scores and transmission behavior of the two feline retroviruses, FIV and FeLV (Table 3 and Figure 4). As indicated in Table 1, FeLV can be spread by a casual contact [46, 47], whereas FIV is predominantly transmitted via non-casual contacts such as bites and cat fights, which are often associated with bleeding and blood contact. Even though sexual transmission is less common, it definitely another mode of transmission [33, 39].

If the proposed shell hardness-PID model is correct then one should expect to see important differences in the shells of these two viruses. Figure 4 shows that FeLV possesses the relatively ordered capsid (34.3$\pm$4.5) and matrix (31.4$\pm$1.7, which are comparable to those of MLV (Matrix PID:37.5$\pm$5.5) and RSV (Matrix PID: 35$\pm$0.1), whereas FIV (Matrix PID:54.3$\pm$1.2) has greater resemblance to the disorder patterns and transmission modes of HIV-2 (Matrix PID: 51.5$\pm$2.5), especially with regard to their capsid disorder.

Moloney MLV vs. Friend MLV shell disorder: Links to oral-fecal transmission

Figure 4 also represents a comparative analysis of two strains of MLV, Moloney MLV (MoMLV) and Friend MLV (FrMLV). It has been experimentally shown that MoMLV spreads more easily via fecal matters and saliva than its counterparts including FrMLV [48]. In agreement with this difference in transmission modes, Figure 4 shows the greater hardness in the matrix of a virus that has greater contribution from the oral-fecal/urine transmission; i.e., MoMLV. One has to be reminded that while the MLV virus has not been detected in large quantities in feces, it is

known to be transmitted via beddings contaminated by saliva and urine especially the urine of male mice[48].

Nucleocapsid and matrix disorder: Operational vs. protective roles

Similar to the other non-primate retroviruses, both feline viruses have high nucleocapsid PIDs. As mentioned above, this has to do with the dual roles that the RNA-binding protein plays. It needs at least a certain level of disorder to play its operational roles, such as coordination of the reverse transcription and participation in evading of the immune system if it could, but yet plays virion-protective roles, if necessary. This is the reason that we tend to see higher disorder at the nucleocapsid when the capsid and matrix are more ordered. Conversely, the nucleocapsid is likely to take over some of the protective roles when the capsid and matrix are highly disordered as we see in HIV-1.

Figure 5A illustrates how the highly flexible RSV nucleocapsid binds to the grooves of the RNA. Figure 5B, on the other hand, shows a somewhat more rigid HIV-1 nucleocapsid with smaller disordered regions in the approximately same areas as the disordered regions in RSV nucleocapsid. This emphasizes the functional importance of some disordered regions. Additionally, our BLAST study shows significant alignment between the protein sequences of the nucleocapsid proteins of RSV and HSV for the portions shown in Figure 5. It is remarkable that despite the high sequence similarity (37% for 2ihx.pdb and 2m3z.pdb), large differences in disorder can also be found (PID: 68.5$\pm$10.5 Vs. 39.5$\pm$3.0, see Figure 5).

Further evidence of the protective roles has already been mentioned above and can be found in statistical analyses seen in Figure 3. While non-primate retroviruses tend to have harder and ordered outer shells (M PID Oneway ANOVA: $p<0.01$, $F = 48.5$) to protect their virions as a result of greater oral-fecal activities, these viruses, unlike the primate ones, have also inner shells

of surprisingly high disorder (N PID Oneway ANOVA: $p < 0.01$, $F = 76.7$). This disorder trend can be noticed in Figures 1-4. The trend demonstrates that greater protective roles (i.e. more rigid and less disordered shell) are given to another layer when one layer is more disordered in an effort to evade the immune systems. Obviously, the outer layers are apparently more imperative when the virus is likelier to be exposed to harsher environments.

Larger disorder at the capsid of HIV-1

Figure 6 illustrates crucial disorder differences among the various retroviruses. The large areas of capsid disorder in red (Figure 6A) can be seen for the HIV-1, in contrast to the less disordered capsid proteins of HIV-2, HTLV-1, RSV and MLV.

Human (HTLV and HIV) retroviruses have more disordered matrix proteins: Less oral-fecal transmission and more immune evasion

While capsid PID levels of some human retroviruses are similar to those of non-primates, the differences between the shells of these viruses become more obvious when their matrix PID levels are compared. Figure 7 shows the qualitative differences in disorder. The disordered regions denoted in red seen in HTLV-1 matrix (Figure 7A) are typically seen in many primate retroviruses, with the exception of many strains of HIV-1, which are even more disordered. The remarkable lack of disorder in non-primate retroviruses is likely an indication of higher levels of oral-fecal transmission component (Figure 7B).

One would also notice that the MoMLV used for Figure 7A has a highly ordered matrix, whereas FrMLV, on the other hand, has a much higher matrix PID (43%), which is surprisingly closer to that of HTLV-1 (41%). It should therefore not to be assumed that all non-primate retroviruses must have matrix proteins that are highly ordered or that all non-primate retroviruses must have an oral-fecal transmission. This was the case, when it was experimentally shown that

only MoMLV was likely to have a noticeable level of oral-fecal transmission [48]. This is highly consistent with our data.

Discussion

Useful tool for epidemiological studies

We have seen that there is evidence that the patterns of disorder at the viral shell varies with the transmission behavior of the retrovirus. For this reason, the PID analyses done here can be a useful tool for epidemiological studies of retroviruses. It allows the investigators to look for clues in behaviors that are hinted by the model.

Hints of effective immune evasion by HIV-1: Virulence and unique disorder pattern

Another implication of our findings is the potential relationship between the shell disorder and immune evasion. The search for an HIV-1 vaccine has been a major project that has yet to show success, even though effective vaccines for its cousin, EIAV, have been found a long time ago[31, 32]. It is therefore not coincidental that we are seeing large differences in disorder for the two viruses at all shell levels. Another hint of immune evasion lies in the differences in virulence among the various retroviruses, such as closely related HIV-1 and HIV-2. Within 10 years, the HIV-1 will kill virtually all of its human hosts who are without anti-viral therapy, whereas HIV-2 usually requires a longer period[1, 30, 37]. The differences can be traced to their viral loads, which can be found at higher levels for HIV-1 and are, as argued in this paper, the result of HIV-1 effectiveness in evading the immune system.

A further hint of evidence for immune evasion lies in the unexplained pattern that has been observed by epidemiologists for a long time: The more sexually promiscuous a population is, the more virulent, the HIV-1 strain tends to become[1, 30]. The mechanisms of immune escape as

describe in this paper could provide a lead towards solving this puzzle.

Possible mechanisms for immune evasion using shell protein disorder

There are various possible mechanisms for immune evasion. A first mechanism involves weak binding of antibodies and related molecules to virus or viral proteins. This is likely to involve some coordination with the surface glycoproteins [1, 49]. While disordered proteins can cause weak binding of antibodies and similar molecules, protein intrinsic disorder can also cause promiscuous binding [7, 50]. Such will allow the virus move into hard to reach places and latently hide there. A third way has to do with the speed in the way the virus replicates. In the case of HIV-1, the rapid replication [4] may originate from the relatively high levels of disorder in all three Gag proteins, which are crucial in the quick virion assembly and other processes important in replication. Its high viral load will help defeat the immune system's attempts to eliminate the virus. Further proofs that highly disordered outer shells provide pathways for more efficient immune escapes can be found in the case of both HIV-1 and the HSV, which, as of to date, have no effective vaccine[8], and, not surprisingly, both viruses have highly disordered matrix proteins [1, 7]. It should be reiterated that despite a search of a large variety of non-related viruses, highly disordered outer shells have yet to be found with the exception of HIV-1 and, to a somewhat lesser extent, HSV[1, 6]. However, some of the mentioned retroviruses such as FIV could also be considered as part of the exceptions as we have seen, even though we must suspect that many of the retroviruses are not be consider as part of the exceptions given their low PID at the matrix.

Further experimental and theoretical bases for immune evasion via shell disorder: Glycoconjugate vaccines

A widely held assumption among many scientists has to do with the belief that the portion of the antigen that matters most is the exposed region since it is this region that will be

recognized by the antibodies. However, a misconception seen is one that is related to the common claim that the buried portion is irrelevant in vaccine development. Such flaw is at odd with some of the fundamentals that have been known about glycoconjugate vaccine development since the 1920s. In 1929, Avery and Goebel found that polysaccharides of specific bacteria used as vaccines are essentially ineffective, but when they are held together as a rigid conjugate via a protein, evidence of improved immunologic memory of the antigen was observed[51, 52]. The exploration of the concept of shell disorder and immune evasion takes this paradigm even further by asking: What happens if the conjugate protein is highly flexible as exemplified by the matrix in HIV-1? Furthermore, a description of how disorder at the shells could increase the motions of the surface glycoprotein has been described in a previous paper[1] while keeping in mind that the HIV-1 is one of the most heavily glycosylated viruses known[49].

Hints for interpreting disorder at various shell levels

There is still a lingering puzzle related to the interpretation of shell disorder. As we have seen, EIAV is unique in our sample of retroviruses by having highly ordered shells at all level. This is in sharp contrast to the retroviruses that are suspected to have higher oral-fecal components such as FeLV and MLV, which have ordered outer shells but yet have higher disorder at the nucleocapsid. Incidentally, the characteristic of having ordered proteins at all shell levels seems also somewhat contradictory to our unpublished data on shell proteins of many non-retroviral arboviruses. The non-primate retroviruses, like FeLV and MLV, tend also to have high disorder at the inner shell, nucleocapsid, but yet have a somewhat more ordered outer shells. How can we interpret the significance of such enigmatic trend in terms of immune evasion?

Inner Vs Outer shell disorder: Latency and antibody escape

A hint of the answer can be found when we look at the viral shell disorder and the nature of the behaviors of the various viruses and their hosts of concern. For instance, EIAV, unlike most of the other arboviruses, does not replicate within its insect vector[33], which acts simply as a carrier. This means that the EIAV has just to deal with the immune system of its mammalian host without having to protect itself from the immune system of the insect host, unlike many other arboviruses. High disorder in the inner shells reflects on the need to evade the multiple host immune systems by providing the ability to lie in latency in specific organs and hosts as in the case of many arboviruses in their reservoirs. As reminded, an accompanying property of having hard outer shells also protects themselves from the destructive enzymes found in the insect saliva. The mechanisms for viral latency are likely to involve the principle of promiscuous binding of disordered proteins especially to proteins of specialized cells that are hidden from the immune systems as mentioned above. By contrast, a key to understanding of the role of disorder at the outer shell is likely to lie in characteristics and properties of the HIV-1 with its highly disordered matrix. Not surprisingly, the HIV-1 has shown exceptional ability to evade the immune system both by hiding in vital organs such as the brain[9, 30] and by rendering neutralizing antibodies[1, 53] ineffective as described previously. This interpretation therefore reiterates the warning that not all retroviruses are likely to have the full capability of HIV-1 in terms of immune evasion, given the varying shell disorder of the different retroviruses seen in our data.

Further applications: Tumor oncolysis and animal models for vaccines

A better understanding of the correlation between the shell disorder and immune evasion might have multiple applications in cancer and vaccine research. One application is the field of oncolytic virotherapy, which is the use of viruses to “attack” tumors [54, 55]. Since it has been known that the immune system has often render therapeutic viruses ineffective, the results

represented here could help researchers to design or identify viruses that are more efficient in evading the immune systems of their hosts. It should be remembered that while oncogenic retroviruses are likely to be unsuitable as oncolytic agents because of their ability to cause cancer, the approach described in this paper can be used to suggest new strategies for oncolytic cancer therapy even if they involve non-retroviruses. Another application is related to the way animal viruses are used to model human viruses, such as the HIV. This paper reminds researchers that there are important differences in the way non-primate and human viruses evade the immune systems arising from the ways they evolved via transmission modes, and the model used here suggests ways to interpolate and take into account of such factors.

Conflict of Interests

GG is an independent researcher and the owner of Goh's BioComputing, Singapore. The authors have no conflict of interests.

Funding

This research is not supported by any external fund.

Figures

Figure 1. PID scores of the gag proteins of some known lentivruses. Viruses ("ALL") in general are compared to the lentiviruses. "ALL" refers the collective viruses available in the database including non-related viruses and non-retroviruses.

Figure 2. Comparison of PID of shell proteins of retroviruses that infect primates.

Figure 3. Human Vs Non-primate animal viruses: PID of shell. Oneway ANOVA has found the mean nucleocapsid PID of the non-primate retroviruses to be statistically higher than that of primate retroviruses. ($p < 0.01$, $F = 76.7$) with the exception of EIAV since EIAV is unusual among retroviruses by having an insect transmission mode. Another Oneway ANOVA using M PID among primates and non-primate retroviruses, excluding FIV and EIAV, also yielded statistically significant results ($p < 0.01$, $F = 48.5$)

Figure 4. Gag PID comparison of MLV strains, feline retroviruses and human retroviruses

Figure 5. Three dimensional representations with disorder annotation. A) Nucleocapsid of RSV B) Nucleocapsid of HIV-1. The regions colored red are areas predicted to be disorder by VLXT.

Figure 6. Three dimensional representations of parts of HIV-1 capsid with other retroviral capsid proteins. A) HIV-1 B) HIV-2 C)HTLV-1 Capsid D) Avian RSV Capsid

Figure 7. Three dimensional representation of portions of primate and non-primate retroviral matrix proteins. A) HTLV-1 matrix B) Moloney MLV (MoMLV) matrix

Tables

Table 1. A selection of a variety of known retroviruses. The viruses are grouped by genus[9], which indicates genetic proximity.

Virus	Genus	Host	Transmission*
EIAV (Equine Infectious Anemia Virus)	Lentivirus	Equines	Insect Vector[33, 34]
FIV (Feline Immunodeficiency Virus)	Lentivirus	Cats	Fights, Bites, Sexual[+]
HIV-1, SIVcpz (Human//Simian Immunodeficiency Virus)	Lentivirus	Human, Monkey	Sexual, Breast Feeding[30]
HIV-2, SIVmac	Lentivirus	Human, Monkey	Bites, Sexual[30, 38]
HTLV-1, STLV-1 (Human//Simian T-Cell Lymphotropic Virus)	Deltaretrovirus	Human, Primate	Breast Feeding, Sexua[4, 30, 41]l
HTLV-2, STLV-2	Deltaretrovirus	Human, Primate	Intravenous Drug Usage[9]
HTLV-3, STLV-3	Deltaretrovirus	Human, Primate	Unknown[43]
MLV (Murine Leukemia Virus)	Gammaretrovirus	Mouse	Sexual, Oral-Fecal/Urine[+], Breast-Feeding[48]
FeLV (Feline Leukemia Virus)	Gammaretrovirus	Cats	Casual Contacts[33, 47]
RSV (Rous Sarcoma Virus)	Alpharetrovirus	Birds	Contacts,[33, 46]

*Column does not attempt to provide an exhaustive list of transmission modes for each virus but, rather, it tries to identify the most important modes of transmission for a given virus.

[+]While sexual transmission may be a less common form of transmission, in FIV, it is not clear if it is or is not a an important mode of transmission. Also, even though MLV virus has not been detected in the feces of mice, it has been detected in urine and saliva.

Table 2 Sample UniProt and PDB accession codes for the proteins. MLV is listed as two known strains, MoMLV (Moloney MLV) and FrMLV (Friend MLV). A much larger database of UniProt accession codes used is available upon request.

Virus	GAG UniProt[+] Accession	PDB Accession Nucleocapsid(N)	PDB Accession Capsid (C)*	PDB Accession Matrix (M)*
EIAV	P69732	2BL6 (p11)	2EIA (p26)	1HEK (p15)
FIV	P16087	(p13)	(p22)	(p15)
HIV-1, SIVcpz	P03348	2M3Z (p7)	1AFV (p24)	1HIW (p17)
HIV-2, SIVmac	P04584	2EC7 (p7)	2WLV (p24)	2K4H (p17)
HTLV-1, STLV-1	P03345	1G03 (p15)	(P24)	1JVR (p19)
HTLV-2, STLV-2	P03346	(p15)	(p24)	(p19)
HTLV-3, STLV-3	Q095Z9	(p15)	(p24)	(p19)
MoMLV	P03332	(p10)	1U7K (p30)	1MN8 (p15)
FrMLV	P26806	(p10)	(p30)	(p15)
FeLV	P10262	(p10)	(p30)	(p15)
RSV	P03354	2IHX (p12)	1EM9 (p27)	(p19)

*Not all of the proteins mentioned can be found in PDB.

+ http://www.uniprot.org/

Table 3. PID levels of shell proteins in various retroviruses.

Virus	PID of Nucleocapsid (N)	PID of Capsid (C)	PID of Matrix (M)	Sample Size[+]
EIAV	26±0.1*	29±0.1*	13±0.1*	3
FIV	71.3±0.5	37.2±0.5	54.3±1.2	3
HIV-1, SIVcpz	39.5±3.0	44.5±2.6	56.5±10.8	12
HIV-2, SIVmac	46.5±9.8	26.6±2.9	51.5±2.5	4
HTLV-1, STLV-1	49±0.1	35±0.1	41±0.1	2
HTLV-2, STLV-2	51±0.1	22±0.1	46±0;1	1
HTLV-3, STLV-3	36.5±0.1	31±0.1	45±0.1	2
MLV	81.5±1.5	46.5±0.5	37.5±5.5	2
FeLV	73.3±5.4	34.3±4.5	31.3±1.7	3
RSV	68.5±10.5	47±2	35±0.1	2

*The standard error is denoted by the prefix, "±"

[+] The number of samples used for each of the GAG protein

Figures

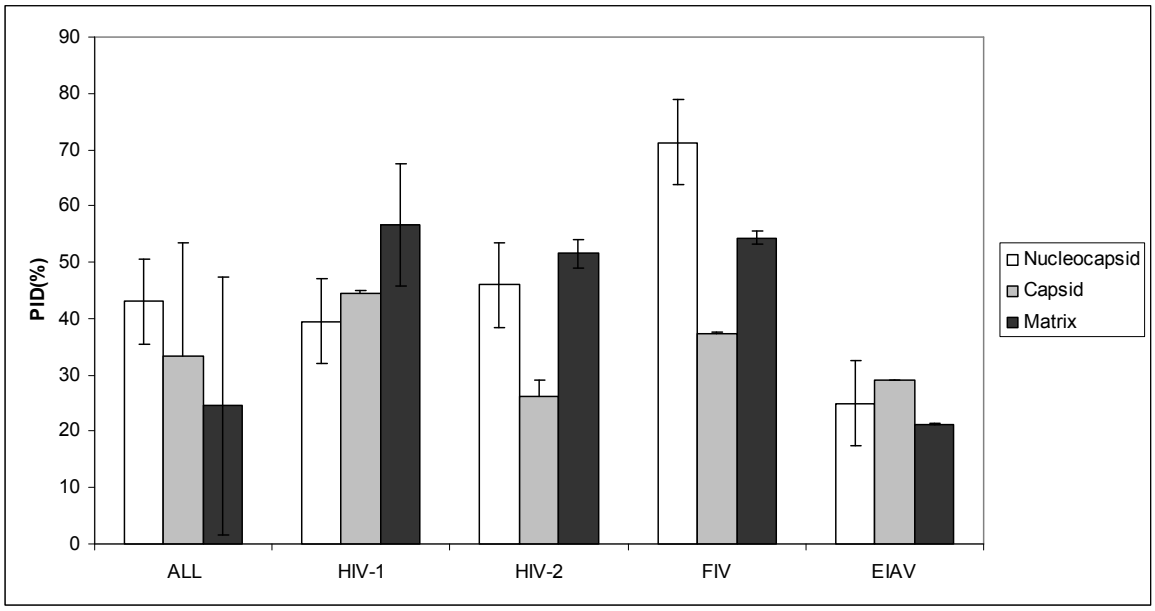


Figure 1.

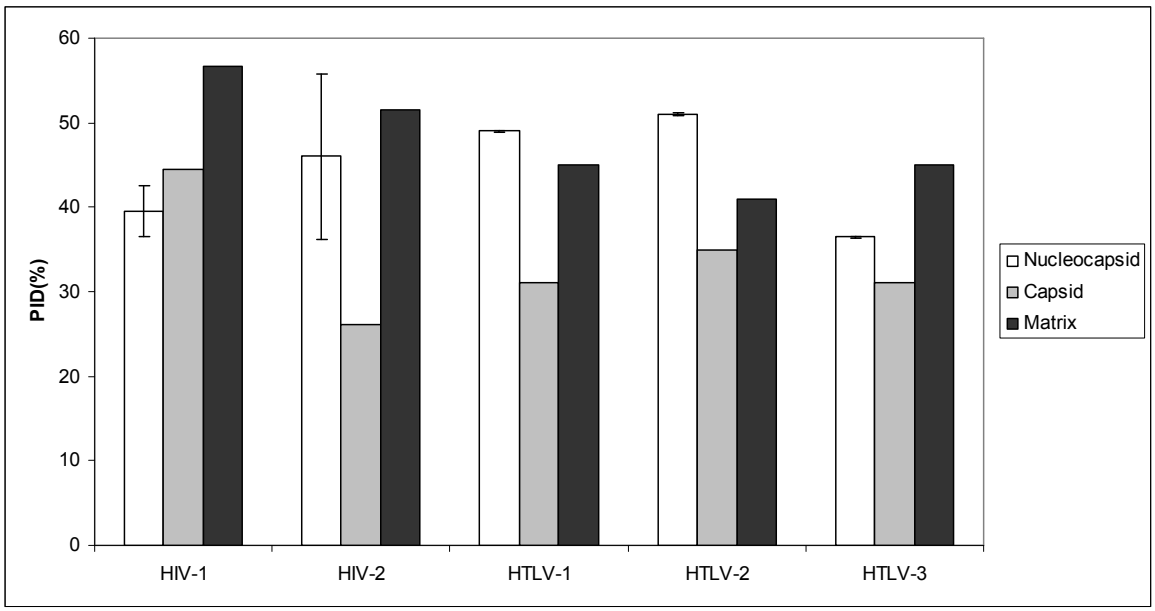


Figure 2.

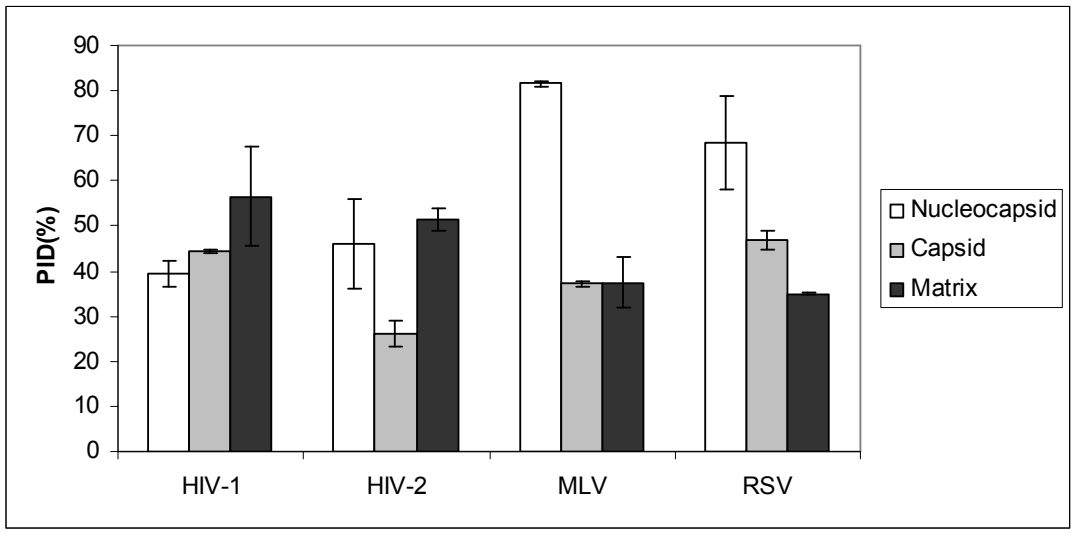
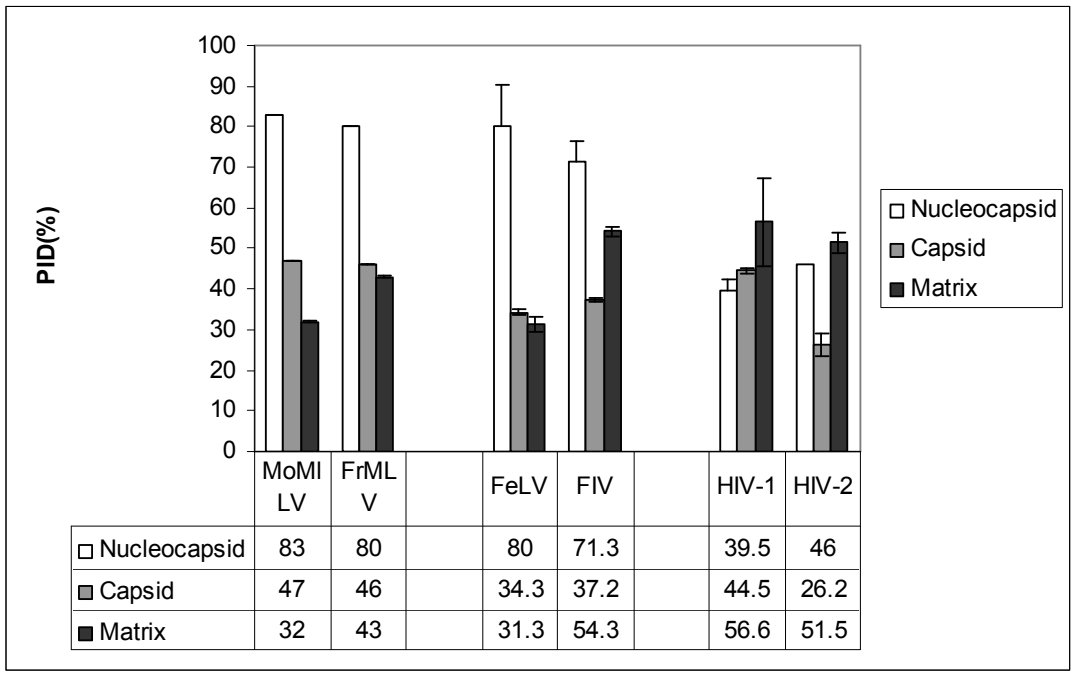


Figure 3.



	MoMLV	FrMLV		FeLV	FIV		HIV-1	HIV-2
Nucleocapsid	83	80		80	71.3		39.5	46
Capsid	47	46		34.3	37.2		44.5	26.2
Matrix	32	43		31.3	54.3		56.6	51.5

Figure 4.

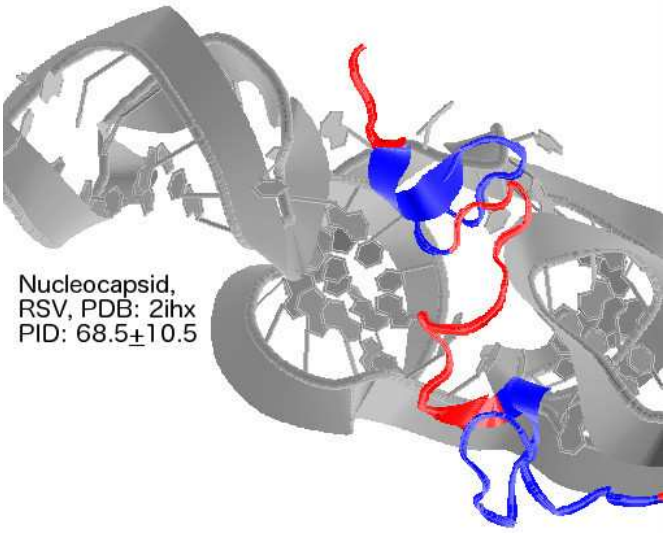


Figure 5A

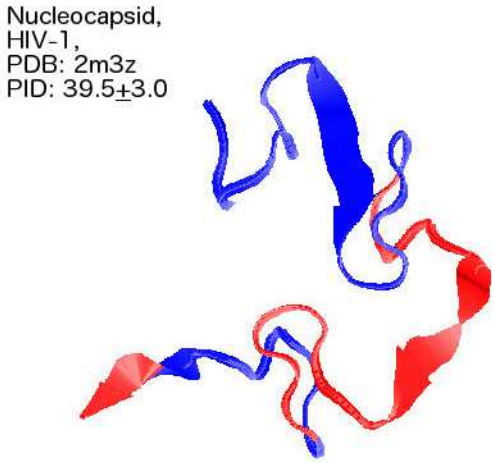


Figure 5.B.

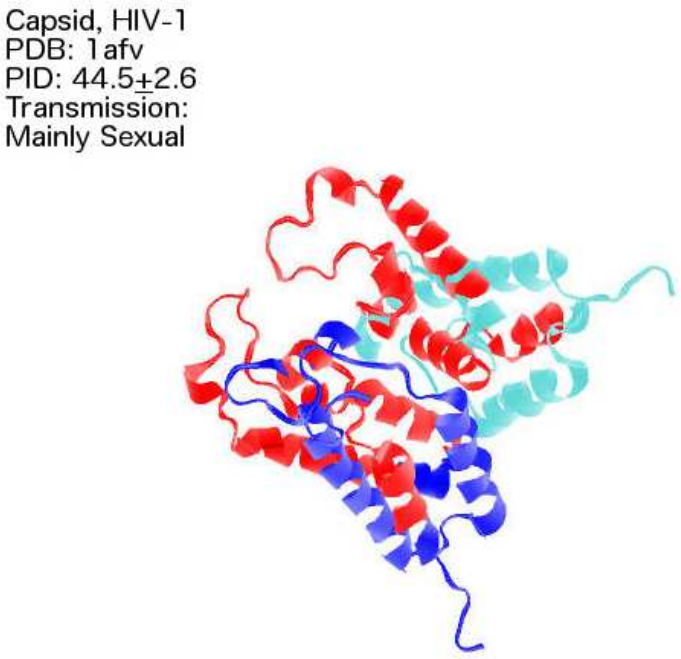


Figure 6A

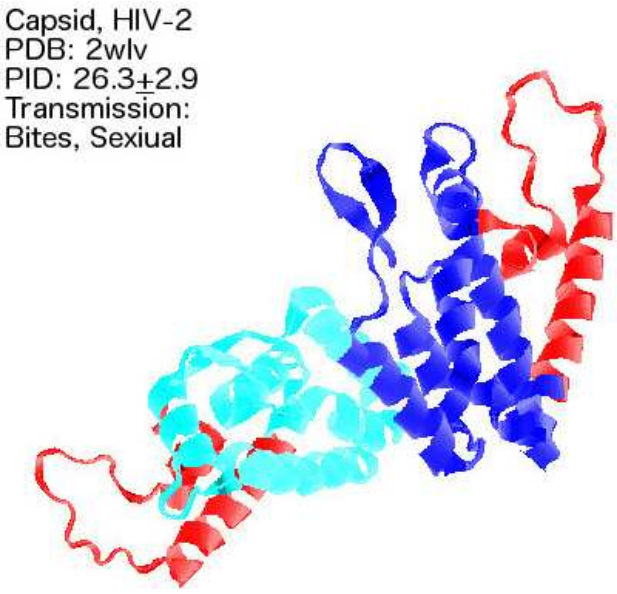


Figure 6B

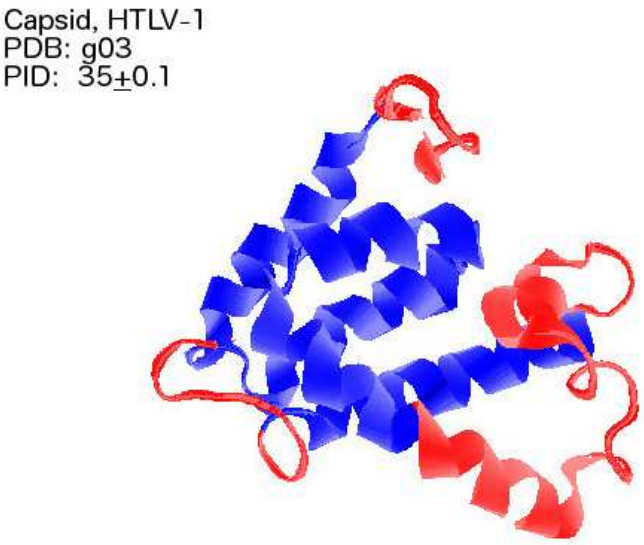


Figure 6C

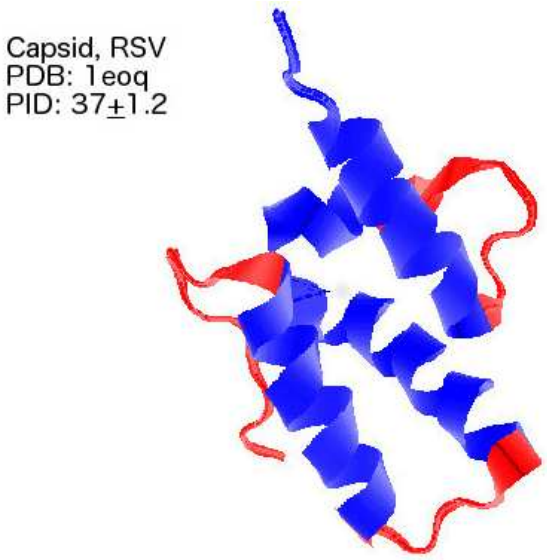


Figure 6D

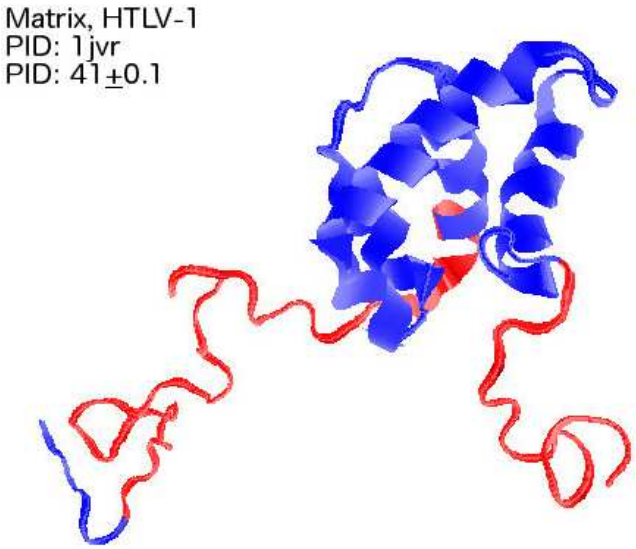


Figure 7A

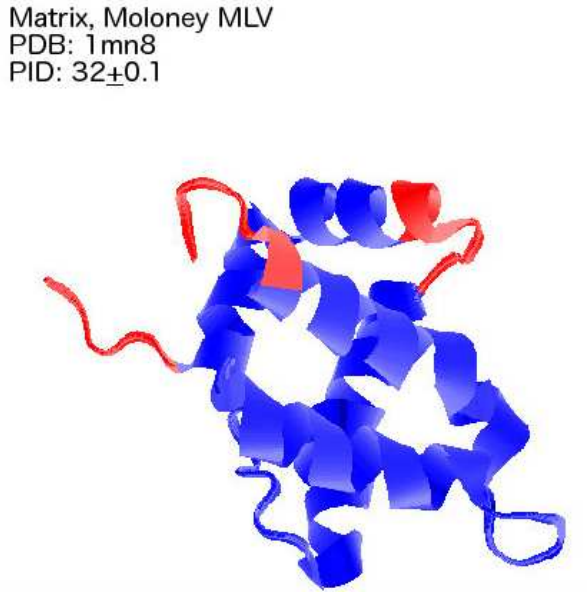


Figure 7B.

References

1. G. K. M. Goh, V. Uversky and A. K. Dunker, *Virol. J.*, 2008, **5**, 126.
2. G. Goh, A. Dunker and V. Uversky, *J Pathog*, 2012, **2012**, 738590.
3. G. Goh, A. Dunker and V. Uversky, *PLOS Currents Outbreak*, 2013.
4. L. Somapyrac, *How pathogenic viruses work*, Jones & Bartlett Learning, 2002.
5. B. Xue, R. Williams, C. Oldfield, G. Goh, A. Dunker and V. Uversky, *Protein Pept Lett.*, 2010, **17**, 932-951.
6. G. K. M. Goh, A. K. Dunker and V. Uversky, *BMC Genomic*, 2008, **9**, S4.
7. A. J. Fonseca, M. Heinemann, R. Leite and J. Reis, *Arch Virol.*, 2010, **55**, 1687-1692.
8. J. Cohen, *Science*, 2010, **330**, 304.
9. N. Acheson, *Fundamental of molecular virology*, Wiley, 2007.
10. P. Romero, Z. Obradovic, C. R. Kissinger, J. E. Villafraca, E. C. Garner, S. Guilliot and A. K. Dunker, *Pac Symp Biocomp*, 1998, 437-448.
11. E. Garner, P. Romero, A. K. Dunker, C. Brown and Z. Obradovic, *Genome Informatics*, 1999, **10**, 41-50.
12. G. K. Goh, A. K. Dunker and V. N. Uversky, *Virol. J*, 2009, **6**, 69.
13. I. Eckerle, M. Müller, S. Kallies, D. Gotthardt and C. Drosten, *Virol J*, 2013, **10**, 359.
14. N. Ferguson and M. Van Kerkhove, *Lancet Infect Dis.*, 2014, **14**, 93-94.
15. V. Uversky and A. Dunker, *Biochim Biophys Acta.*, 2010, **1804**, 1231-1264.
16. H. M. Berman, Z. Westbrook, G. Feng, H. Bhat, H. Weissig, L. N. Sindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235-242.
17. X. Li, P. Romero, M. Rani, A. K. Dunker and Z. Obradovic, *Genome Inform Ser Workshop Genome Inform*, 1999, **10**, 30-40.
18. X. Deng, J. Eickholt and J. Cheng, *Mol Biosyst*, 2012, **8**, 114-121.
19. B. He, K. Wang, Y. Liu, B. Xue, V. Uversky and A. Dunker, *Cell Res*, 2009, **19**, 1929-1949.
20. J. Bourhis, B. Canard and S. Longhi, *Curr Protein Pept Sci.*, 2007, **8**, 135-149.
21. F. Ferron, S. Longhi, B. Canard and D. Karlin, *Proteins*, 2006, **65**, 1-14.
22. V. Uversky, *Protein Scince*, 2013, **22**, 693-724.
23. P. Tompa, *Trends Biochem Sci*, 2002, **27**, 527-533.
24. J. Nadeau, *Introduction to experimental biophysics: Biological methods for physical scientists*, CRC Press, 2012.
25. P. E. Wright and H. J. Dyson, *J Mol Biol*, 1999, **293**, 321-331.
26. J. Fong, r. B. Shoemake, S. Garbuzynskiy, M. Lobanov, O. Galzitskaya and A. Panchenko, *PLoS Comput Biol*, 2009, **5**, e1000316.
27. A. Herráez, *Biochemistry and Molecular Biology Education*, 2006, **34**, 255-261.
28. I. Korf, M. Yandell and J. Bedell, *BLAST*, O'Reilly, 2003.
29. J. Goudsmit, *Viral Fitness: The Next SARS and West Nile in the Making*, Oxford University Press, Oxford, 2004.
30. J. Goudsmit, *Viral Sex: The Nature of AIDS*, Oxford University Press, New York, 1997.
31. J. Du, X. Wang, J. Ma, J. Wang, Y. Qin, C. Zhu, F. Liu, Y. Shao, J. Zhou, W. Qiao and X. Liu, *Retrovirology*, 2014, **11**, 26.
32. T. Tagmyer, J. Craigo, S. Cook, D. Even, C. Issel and R. Montelaro, *J. Virol*, 2008, **82**, 4052-4063.
33. S. Aiello and A. Mays, *Merck Veterinary Manual*, 8th edn., Merck & Co, 1998.
34. F. Baldacchino, V. Muenworn, M. Desquesnes, F. Desoli, T. Charoenviriyaphap and G. Duvallet, *Parasite*, 2013, **20**, 26.
35. D. Malamud, W. Abrams, C. Barber, D. Weissman, M. Rehtanz and E. Golub, 2011, **23**, 34-37.

36. P. A. Marx, Y. Li, N. W. Lerche, S. Sutjipto, A. Gettie, J. A. Yee, B. H. Brotman, A. M. Prince, A. Hanson and R. G. Webster, *J Virol.*, 1991, **65**, 4480-4485.
37. S. Nyamweya, A. Hegedus, A. Jaye, S. Rowland-Jones, K. Flanagan and D. Macallan, *Rev Med Virol.*, 2013, **23**, 221-240.
38. S. Locatelli and M. Peeters, *AIDS*, 2012, **26**, 659-673.
39. J. Yamamoto, M. Sanou, J. Abbott and J. Coleman, *Curr HIV Res.*, 2010, **8**, 14-25.
40. A. Deshpande, S. Jadhav and A. Bandivdekar, *Aids Res Ther*, 2011, **8**, 16.
41. R. Gallo, *Best Pract Res Clin Haematol.*, 2011, **24**, 559-565.
42. N. Rosenberg and P. Jolicoeur, in *Retroviruses.*, ed. H. S. Coffin JM, Varmus HE, editors., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, New York, 1997.
43. A. Gessain, R. Rua, E. Betsem, J. Turpin and R. Mahieux, *Virology*, 2013, **435**, 187-199.
44. R. Mahieux, J. Pecon-Slattery, G. Chen and A. Gessain, *Virology*, 1998, **25**, 71-84.
45. M. Phillips, *Environ Health Perspect.*, 2005, **113**, A810.
46. B. Burmester and T. Fredrickson, *Avian Diseases*, 1966, **10**, 259-287.
47. B. Willett and M. Hosie, *Vet J.*, 2013, **195**, 16-23.
48. R. Buffett, J. Grace, L. DiBerardino and E. Mirand, *Cancer Res*, 1969, **29**, 588-595.
49. D. J. Vigerust and V. L. Shepherd, *Trends Microbiol.*, 2007, **15**:211-218.
50. A. Dunker, M. Cortese, P. Romero, L. Iakoucheva and V. Uversky, *FEBS*, 2005, **272**, 5129-5148.
51. O. Avery and W. Goebel, *J. Exp. Med*, 1931, **154**, 437-447.
52. D. Goldblatt, *Clin Exp Immunol.*, 2000, **119**, 1-3.
53. D. R. Burton, *Nat Rev Immunol.*, 2002, **2**, 706-713.
54. Y. Chernajovsky, L. Layward and N. Lemoine, *BMJ*, 2006, **332**, 170-172.
55. M. Ferguson, N. Lemoine and Y. Wang, *Adv Virol.*, 2012, **2012**, 805629.