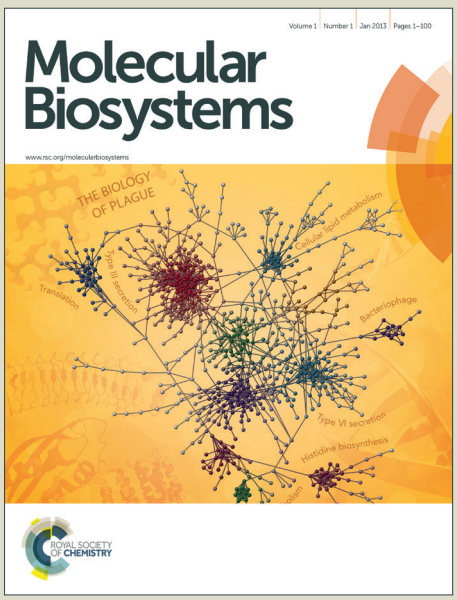
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Detection of links between Ebola nucleocapsid and virulence using disorder analysis

Gerard Kian-Meng Goh (gohsBioComputing@yahoo.com)[1], A. Keith Dunker[2], Vladimir N. Uversky[3,4,5,6]

1. Goh's BioComputing, Singapore, Republic of Singapore

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

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

4. Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow region, Russia

5. King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia;

6. Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia

Key words: intrinsically disordered protein; nucleocapsid; Ebola; virulence; viral protein; protein structure; protein function

Abstract

The underlying reasons for the differences in the virulence of various types of Ebola virus (EBOV) remain unknown. Comparison of the percentage of disorder (PID) in the nucleocapsid proteins VP30 and NP reveals high correlation between nucleocapsid PIDs and the case-fatality rates of EBOV. The higher disorder of these proteins is likely to be needed for more efficient multiplication of virus copies via more efficient viral RNA transcription and the more promiscuous protein binding potential. This is important for the more efficient assistance of nucleocapsid in viral particle budding and of the assembly and mobility of viral proteins across host membrane and within the cytoplasm. A more comprehensive knowledge of the molecular mechanisms of EBOV virulence would also lead to new and more effective strategies in vaccine development

Introduction

Marburg virus (MARV) and Ebola virus (EBOV) are enveloped single-stranded negative-sense non-segmented RNA viruses from the *Ebolavirus* and *Marburgvirus* genera of the *Filoviridae* family from the *Mononegavirales* order. These viruses are known to cause severe hemorrhagic fever in humans and other mammals that is usually fatal[1]. While EBOV and MARV are of different genii, they have much in common. The RNA genomes of both viruses are similarly non-segmented and are of the same size, 19 kb[2]. More interestingly, both EBOV and MARV produce four proteins from their nucleocapsid genes, not three as in most of the other filoviruses[3]. There are, however, some differences. One of these has to do with the fact that MARV and EBOV replicate using different strategies as we shall see later.

There are five EBOV types that have been identified since this virus was first discovered in 1976, when there were two simultaneous outbreaks in Africa caused by two different EBOVs. They were named the Zaire EBOV (ZEBOV) and Sudan EBOV (SEBOV) and had case-fatality rates (CFRs) of 88% and 53%, respectively [2, 4]. An EBOV strain Reston EBOV considered as non-pathogenic to human was found among laboratory monkeys imported from the Philippines in 1989 [5]. In the most recent Ebola outbreak of 2013-2014 (as of October 8, 2014), a total of 8,399 confirmed, probable, and suspected cases of Ebola virus disease (EVD) have been reported in seven affected countries (Guinea, Liberia, Nigeria, Senegal, Sierra Leone, Spain, and the United States of America). There have been 4,033 deaths [6]. The current EVD epidemic is the most severe outbreak since discovery of Ebola, since EVD cases from this single outbreak exceeded the sum of all previously identified cases [7]. The MARV was first discovered in 1967, when infected monkeys transmitted the virus to laboratory workers in Marburg, Germany. The mortality rates of MARV virus can reach over 80% [8]. While various EBOV species have different levels of virulence, the factors contributing to the virulence remain poorly understood[5]. This is mainly because the research on filoviruses has been hampered by their status as biosafety level 4 pathogens.

Many biologically active proteins are intrinsically disordered; i.e., characterized by the lack of stable, well-defined structures[9, 10]. Comparison of ordered and intrinsically disordered proteins (IDPs) revealed that their amino acid sequences are noticeably different, that IDPs and IDP regions (IDPRs) share at least some common sequence features over many proteins, and that IDPs/IDPRs can be reliably predicted based on the amino acid sequence alone[11-14] using various disorder predictors, such as PONDR® VLXT [15, 16].

IDPs/IDPRs are common in viruses, where they play a number of crucial roles [17]. The concept of protein intrinsic disorder has been successfully used to study virulence, evolution, and transmission mode of various viruses, such as HIV, influenza A 1918 H1N1 and H5N1 viruses, SARS-CoV, MERS-CoV, HPV, and HCV[18-26] (17-24). For example, proteins of the HIV-1 outer shell and matrix were found to possess high disorder levels[18, 26], and this abundance of disorder was suggested to contribute to the virus innate ability to evade the host immune system with great effectiveness [18]. The grouping of the coronaviruses (CoVs) by the disorder levels of their shell proteins generated a model that was able to predict moderate respiratory and oral-fecal modes of transmission for the SARS-CoV [20] and lower respiratory and higher oral fecal transmission potentials of the MERS-CoV [21].

EBOV genome codes for nine proteins: RNA-directed RNA polymerase L, nucleoprotein (NP), envelope glycoprotein (GP), membrane-associated protein VP24, minor nucleoprotein VP30, polymerase cofactor VP35, matrix protein VP40, pre-small/secreted glycoprotein (sGP), and super small secreted glycoprotein (SsGP). The goal of our research was to address the issue of virulence variability by implementing the comparative computational analysis of the levels of intrinsic disorder found in the EBOV/MARV nucleocapsid proteins NP (nucleoprotein) and VP30 (minor nucleoprotein).

Material and Methods

Tools

An important computational tool used in this study is PONDR® VLXT (www.pondr.com), which predicts regions of disorder given the input sequence of amino acids from a protein [15, 16]. To simplify the use of this tool, relational database for automatic storage of protein information by sequence,

disorder prediction and virus types was developed using MySQL and JAVA [18, 26]. Using JAVA and Jmol, graphical tools were also designed for generation of 3D structures of proteins with disorder annotations (18-20). Furthermore, an automated data entry system was developed in JAVA for links to NCBI-PDB (http://www.ncbi.nlm.nih.gov/Structure/index.shtml) and UniProt (http://www.uniprot.org). The JAVA programs and MySQL databases were built to provide solutions as seen in the mentioned previous papers..

PONDR® VLXT was also used to generate percentage of intrinsic disorder (PID), which is the number of residues found to be disordered divided by the total number of residues in a given protein. The R-statistical package was used to Oneway analysis of variance (ANOVA), Multivariate ANOVA (MANOVA) and regression analysis[27]. The standard errors and means of the PIDs were calculated using MySQL, whereas the corresponding values for CFR were obtained using EXCEL

Challenges in data collection and different virulence levels among species

As the goal of this study was to correlate the mortality rates to nucleocapsid disorder, reliable data for the EBOV/MARV CFR had to be collected. This task faced several challenges. One of these difficulties is related to the small number of known Ebola outbreaks, and to even smaller number of MARV outbreaks [2, 5]. Furthermore, many of these outbreaks involve small number of cases especially if they were laboratory accidents. Such small sample sizes make the obtained CFR values rather unreliable. An added complication has to do with the EBOV nature that has five currently known types [4]. There are likely to be more EBOV species lurking in the wild that we are unaware of, and this therefore places a further limitation to data collection. To illustrate this point, the Ivory Coast EBOV (Tai Forest) involved only a single case in which a veterinarian became infected during the examination of the carcass of a monkey[4].

As a further illustration of the complications that the rarity of occurrence of EVD places on our research, we can see that, in the case of Bundibugyo EBOV (BEBOV), there is only one known

outbreak[4] and only one sequence each is available for the two proteins that we are concerned with. While this 2007 outbreak has a sample size that is adequately large (149 cases), it exemplifies the further data constraints we face especially with respect to calculations of the values of the standard deviation of both PID and CFR. It should also be reminded that EVD, unlike AIDS, had been a neglected disease that strikes the poorest countries in the world even as EBOV has been known since 1976. One should therefore not be surprised that EBOV data, including protein sequences, are by no means plentiful, unlike those of HIV.

As mentioned, because many of the outbreaks involve only a handful of cases, the resulting CFRs are often unreliable given the small sample sizes. For this reason, only outbreaks that involved samples sizes of over 50 cases were used in the CFR computation. Such a practice is not only necessary but is also justified by the Central Limit Theorem, which encourages larger sample sizes to be used so as to obtain greater accuracy[27]

Protein sequence selection

Despite the challenges, a limited but adequate number of samples (UniProt accession codes) are available for our study. The main focus of this study is centered on two nucleocapsid proteins, the VP30 (minor) and NP (major) since these are the only two proteins that have positive correlation to virulence as we shall see in this paper. A total of 16 sequences were used for the four EBOV species and two proteins studied. Another four sequences of the MARV nucleocapsid were added to the statistical studies.

Results

Description of the species and outbreaks

As mentioned above, the EBOV was first discovered in 1976, when two outbreaks occurred simultaneously in Zaire and Sudan. The EBOV types involved in these outbreaks were ZEBOV and SEBOV. During the outbreaks, ZEBOV had 318 cases and 284 deaths (CFR-88%), whereas SEBOV involved 284 known infections with 150 deaths (53%) [2, 28]. While there were several outbreaks involving SEBOV, only one involved infections of over 50 people. This was the 2000-2001 Uganda that included 425 cases with the same 53% CFR. For ZEBOV, there were 8 other outbreaks with more than 50 cases each. Like the 1976 outbreak, the ZEBOV outbreaks involved high CFR, usually around 70-90%. There was, however, one ZEBOV outbreak with noticeably lower CFR. This was the 1994 Z EBOV Gabon outbreak, with 52 cases and the CFR of 60% [4]. The REBOV is considered to be non-pathogenic to humans by WHO (World Health Organization), even though some levels of virulence have been detected among non-human primates [5]. Table 1 shows that the PIDs of the corresponding REBOV nucleocapsid proteins remain invariant in different strains in terms of disorder as seen in the low standard errors, given that the analyzed samples are from 1989 and 1996. Another EBOV species, the Bundibugyo EBOV (BEBOV), with lower virulence was discovered in 2007 (Bundibugyo Uganda). This EBOV has a CFR of 25%, which is higher than that of REBOV (0%) but smaller than the SEBOV (53%) [2, 4]. Therefore, in our analysis, we considered all known EBOV types with the exception of the EBOV from the Ivory Coast (Côte d'Ivoire) outbreak of 1994-1995. Although a new subtype of Ebola was isolated (EBOV CI) and although genetic sequence of this virus is readily available, this particular virus subtype was not included in our study since just one person was affected in this outbreak [2, 4].

The MARV is highly virulent with mortality often approaching 90%. However, the known outbreaks often involve small numbers of individuals. A sample that is available and well-characterized is from the Popp-1967 strain, which was the one involved in the laboratory outbreak in West Germany,

where the virus was first isolated. Since then, there were only two MARV outbreaks that involved more than 50 people. The first one involved 154 cases with the CFR of 83% in 1998 (Democratic Republic of Congo, formerly Zaire). The other happened in 2004 Angola outbreaks, which had a CFR of 90% among 252 individuals [2, 8].

Stability in nucleocapsid disorder within EBOV species with a slight anomaly: The 1994 Gabon ZEBOV strain

The PIDs for the VP30 and NP are generally quite stable within an EBOV specie even over a long period of time. For instance, the SEBOV PID values for NP and VP30 remain virtually identical at 42% and 33% respectively for even for distant strains such as those of Boniface 1976 and Uganda 2000. This trend is also seen for ZEBOV with a slight but noticeable exception. Similarly, it can be seen that the PIDs of NP and VP30 remain virtually constant at 43% and 41% respectively even for distant strains such as Mayinga 1976 and Guinea 2014. The mentioned anomaly could, however, be seen for the Gabon 1994 strain, which has an NP PID of 41%. This peculiarity will be revisited in a later section with respect to the lower than usual CFR of the 1994 ZEBOV outbreak in Gabon at 60% as compared to the higher CFR of the other ZEBOV strains that could reach above 80%[5].

PID analysis: Correlation between the nucleocapsid disorder and virulence

The PID offers a scale for measuring the levels of disorder in proteins. This scale is useful in comparative studies involving different proteins or similar proteins across different species. Table 1 shows that there is a statistical difference between the CFRs of various EBOV species. In our previous studies we observed that grouping CoVs by PIDs of their shell proteins generated a prognostic model for prediction of the transmission modes of SARS-CoV [20] and MERS-CoV [21]. Therefore, we hypothesized that the disorder levels in the nucleocapsid proteins of various EBOV

and MARV types correlate with the virulence of these viruses. Table 1 illustrates feasibility of this hypothesis and shows that there is a high correlation between virulence and nucleocapsid disorder. In fact, application of the regression analysis revealed that CFR (dependent variable) and NP/VP30 PIDs (independent variables) data for different species are highly correlated (r^2= 0.92, F=39.4, $p<0.01$). An alternative way is to arrange the virus types by levels of CFR as in Figure 1 and then perform a MANOVA analysis using NP and VP20 PIDs as independent variables with the dependent variable being virus types labeled (0,1,2..etc) according to their CFR ranks as in Figure 1

This MANOVA analysis shows statistically high significance ($p<0.01$, F=27) for the model. These observations are further illustrated by Figure 1 which clearly shows that an increase in the case-death rates is always followed by the PID increase in either minor (VP30) or major (NP) nucleocapsid protein or in both these proteins together.

Consistency of a trend between the virulence and disorder

Figure 1 shows that there is typically rather good correlation between the mortality rates of various virus species and the PID values of their VP30 and NP. This suggests that the PID levels of VP30 and NP can be used as a predictive model for virus mortality. The average CFRs by virus types were calculated, tabulated and shown in Table 1 and Figure 1. The relatively small standard errors of both PIDs and CFRs by virus types should be noted. Furthermore, ANOVA of the mortality rates by virus type does show statistical significance (F=46, $p< 0.01$), which basically tells us that each virus type tends to have its distinct levels of virulence for humans. This fact actually strengthens the MANOVA result mentioned above since the former links EBOV types to virulence and the latter, on the other hand, demonstrates the statistical relationships between nucleocapsid disorder and EBOV species. Another interesting note is that the CFRs of outbreaks with over 50 cases are found to be more consistent and reliable than those with lower number of cases. This characteristic is consistent

with the above-mentioned Central Limit Theorem[27], which predicts that data with larger sample sizes are likelier to have lower margins of error..

Figure 1 also shows that PIDs in both nucleocapsid proteins should be taken into account for virulence evaluation. In fact, while it seems that the increase in the NP PID is the major contributing factor for the difference in virulence between the REBOV (CFR: 0%) and SEBOV (CFR: 53%), the PID of VP30 is an important explanatory factor for the differences in SEBOV and ZEBOV virulence.

Figure 2 represents three-dimentional structures of the VP30 for the non-pathogenic REBOV and highly virulent ZEBOV with disorder annotations and shows that the structures of the VP30 proteins of the two viruses are quite similar with the exception of the presence or absence of disordered retgions. In fact, Figure 2 shows that not only the disordered regions (red) of ZEBOV are much larger in size, but they are more plentiful.

Finally, Figure 3 represents peculiarities of disorder distributions in NP and VP30 of two types of Ebola virus, non-pathogenic REBOV and highly virulent ZEBOV. Since both Ebola types had several outbreaks, several strains were isolated from afflicted patients (see Table 1 for details). Figure 3 shows that for both nucleocapsid proteins, the variation in disorder distribution within a given virus type was much less pronounced than the disorder variability between the types (e.g., compare Figure 3A and 3B or 3C and 3D). It is also clearly seen that nucleoproteins from the virulent Ebola type are noticeably more disordered than the nucleoproteins of the non-pathogenic REBOV.

Statistical Analysis of disorder reaffirms differences in MARV and EBOV transcription.

A quick glance at the VP30 in Figure 1 and Table 1 could lead us to suspect that something is amiss since MARV VP30 PID falls to a low of 32% while MARV are usually highly fatal with CFR reaching well over 80%. In fact, our additional regression analysis using only VP30 PID and CFR as variables shows poor correlation and statistical significance (,r^2 =0.1, p >0.1)). On the other hand, a regression analysis using NP PID and CFR without VP30 PID did find a reasonable correlation and statistical significance (,r^2=0.64,F = 14 P < 0.01). As we have seen, when the regression analysis was made with NP, VP30 and CFR, however, a strong correlation was obtained (r^2 = 95, F =40.2, p < 0.01).

How is this even mathematically possible for VP30 PID to be obviously an important independent co-variable when it performs statistically poorly as a sole independent variable? As we have seen, VP30 PID for MARV is at the low 32% but the corresponding NP PID, on the other hand, reaches 47%, which is a high level not found in any EBOV. With MARV having the highest CFR (> 80), the high NP PID more than fully compensates for the shortfalls of VP30 PID in MARV. This explains the high correlation obtained only when both VP30 and NP PIDs are used as independent variables. The biological reason for this has already been visited: MARV and EBOV have different modes of replication. While VP30 is essential for EBOV transcription, MARV uses NP, not VP30[3]. This not only explains the peculiarity of the data but to some extent, also serves to experimentally validate some of our results.

Discussion

Reasons for initial interest in NP and VP30

The authors first became interested in the shell proteins because in the past we have uncovered peculiar behaviors pertaining to the shell proteins. We have seen also in the case of coronaviruses,

the disorder at matrix and nucleocapsid has implication on the way the viruses are transmitted[20, 21] Furthermore, the matrix protein of the highly virulent HIV-1 was found to be highly disordered especially when compared to the less virulent HIV-2[26]. The matrix of HIV-2 can be found to be less disordered than HIV-1. It has been believed that such higher is associated with the ability to evade the host immune system. In the case of EBOV and MARV, positive correlation with virulence could only be found for NP and VP30.

Potential for unification of all current knowledge of molecular determinant of EBOV virulence

While the molecular determinant of an EBOV/MARV virulence remains unclear, we do suspect where the virulence is arising from. For instance, the bleeding arising from EBOV/MARV infections can be traced to the surface glycoprotein (GP) that can damage the host cell membrane [29]. The GP is therefore one of the primarily suspected determinant of virulence for EBOV and MARV.

Given that the Ebola GP is considered as a determinant of virulence, the question then arises on how this observation is correlated with the findings of our paper. It is known that the virulence arising from GP is dependent on the GP levels in the body [30]. In other words, the viral load matters, which make an important link to the functions of the nucleocapsid proteins that are experimentally known to affect levels of virus in the host. While many of their functions are not well understood, both the minor and major nucleocapsid proteins are directly involved in RNA transcription upon entry into the host cells [31]. Other experimentally verified functions of the nucleocapsid proteins include assembly of the nucleocapsid (together with VP24 and VP35) as well as assistance in the budding of the viral particles and in the mobility of viral proteins across the host membrane and within the cytoplasm[3, 31-34]. The VP30 is an important player during initiation of the EBOV transcription since this protein prevents a premature termination of the process. The NP encapsulates the viral genome and protects it from nucleases. The encapsulated genomic RNA serves as template for transcription

and replication. During replication, encapsulation by NP is coupled to RNA synthesis and all replicative products are resistant to nucleases.

Greater Disorder: More efficient transcription

It is known that RNA- and DNA-binding proteins are typically rather disordered and these high levels of intrinsic disorder are needed for the various processes taking place during transcription and replication [35]. It is feasible that greater disorder can provide greater efficiency during transcription and thus, results in more virus copies, which may be essential for GP abundance and greater virulence of the Ebola virus. In light of these observations, the comparison of disorder data for EBOV and MARV provides some interesting hints. Unlike PIDs of the EBOV VP30s, which range from 33% in the non-pathogenic REBOV to 43% in the highly pathogenic ZEBOV, the PID of the VP30 of the most virulent MARV strain is just 33%. However, the PID levels of the MARV NPs (44% and 47%) are very high, noticeably exceeding the corresponding values found in various EBOVs. Therefore, low PIDs of the MARV VP30s are compensated by high PIDs of their NPs. Curiously, in line with these observations are differences in functional roles of NP and VP30 in EBOV and MARV: VP30 is essential for the EBOV transcription but not for MARV, which relies more on NP [3].

Advantages of promiscuous binding

Another advantage of disorder proteins over ordered proteins is the ability of IDPs for promiscuous binding [36, 37]. Such binding promiscuity is advantageous for viruses too providing crucial means for efficient invasion of various cell types. In fact, the greater virulence is defined by greater ability of a virus to damage more vital organs, and this capability could be determined by the increase binding promiscuity of viral proteins. Therefore, the higher binding promiscuity arising from the greater

disorder levels can help the two other functions of the EBOV/MARV nucleocapsid, such the assistance in the transportation of viral particles across the host membrane and within the cytoplasm prior to budding[32-34]. Both these functions depend on the interactions between the viral and numerous host proteins which are more efficient for protein with higher disorder levels.

The model presented in our study shows that there is a correlation between the EBOV virus types and virulence. A question then arose: Would this model be able to predict the virulence of an EBOV strain during an outbreak regardless of the EBOV type. Given that there were just a few EBOV and MARV outbreaks, one might argue that the sample sizes are not large enough for the development of an accurate model. However, existing data suggest the model can reliably predict the levels of virulence within the EBOV/MARV type. As mentioned previously, this could be found in the data for ZEBOV 1994 Gabon strain, which has the lowest case-fatality rate (60%) among those with 50 or more infections [4], and which is characterized by the lowest PID level of its NP. (41%)

A summary of the known functions of VP30 and NP

As already mentioned, experiments have provided us with only some understanding of the multifunctional aspects of both NP and VP30. The VP30 acts as a transcription factor in the initiation of transcription, while NP serves to protect the RNA-protein complex from cell proteases by spanning the complex during transcription. It is also known that while VP30 plays a major transcriptional role in EBOV, NP takes over much of the role in the case of MARV. NP is also essential in the budding and assembly of viral particles[3, 33]. Furthermore, NP plays a part in the mobility of viral proteins across the host cytoplasm and membrane[3, 31-34]. As an additional note, the presence of VP30 at specific concentration allows enhanced viral transcriptions but VP30 becomes

an inhibitor if over-expressed. With the summarized known functions of VP30 and NP, we can now address the known structural functions of the proteins with respect to our disorder data.

The known structural functions of VP30/NP and the disorder data

While some knowledge of functional domains of VP30 is available especially for the C- terminus, the atomic structure of the larger NP is not available and its structure-function relationships remain poorly understood, even though the C-terminus of ZEBOV NP has been crystallized[31, 38]. The C-terminus of NP resembles in many way uniquely like a beta grasp[38], which has potentials for protein-protein interactions especially with other proteins. The VP30, on the other hand self-assembles into unique dimeric and oligomeric alpha helices[31]. The regions around C-terminus are believed to bind to NP.

Analysis of the patterns of disorder and what is known about NP and VP30 reveals that the changes in disorder for both proteins are likely to have effects on several factors. We know, for example, that locations 1-451 are responsible for NP-NP interactions and the formation tube-like structures, which provide for greater mobility and budding of viral particles. The region around 451-600 is responsible for NP structure-formation and viral replication. Also, incorporation of NP into virus-like particles is dependent on the last 50 AA sequence of NP[3]. Careful analysis of PONDR(r)-VLXT in Figure 3 A-B tells us that all three regions contain subregions with higher disorder in the case of the highly virulent ZEBOV as compared to the non-pathogenic REBOV. This basically implies that disorder is likely to affect virus and VLP productions multi-functionally.

The VP30 region around locations 94-112 forms a hydrophobic stretch that affects the oligomeric formation[31]. Interestingly, a small spot of disorder, near location 100, can be seen for ZEBOV, but not for REBOV (see Figure 3). This suggests that disorder could play a role in allowing more

efficient self-assembly of VP30. While VP30 is known to bind to NP, a complete structural understanding remains elusive. A hint that disorder may play a role is the fact that most of the disordered regions as seen in red in Figure 2A are those exposed and pointing away from the VP30-VP30 bonds. This could mean that the areas most likely to be in contact with NP are enhanced by having higher disorder as in the case of ZEBOV.

Additionally, it can be seen in Figure 2 that many of the disordered regions are parts of the alpha-helices. Such have been also been observed as features of induced folding and molecular recognition including DNA/RNA recognition in other transcription factors[39]. Being also a transcription factor, one could suspect that, at least, some of the VP30 disordered regions are actually both protein and RNA recognition sites. It is therefore likely that VP30 becomes more efficient as a transcription factor by having greater ability to recognize both RNA and other proteins such as NP.

The 2014 West Africa Outbreak

The nucleocapsid PIDs from the current ZEBOV Guinea 2014 outbreak have been included in our database. The genome used here is of a strain from Gueckedou (H.sapiens-wt/GIN/2014/Gueckedou-C07H) [40]. The nucleocapsid PIDs of this strain (NP PID =43%, VP30 PID = 41%) are not too different from those of the other ZEBOV strains, such as the ZEBOV 1976 that have higher case-fatality ratio. Since the 2014 Guinea outbreak is still ongoing at the time of writing of this paper, the present case-fatality ratio is still changing and unreliable, even though some estimates place it around 70%[41]. While this figure is indicative of a high level of virulence comparable to that of the ZEBOV 1976, large amounts of mutations have been observed as the EBOV is rapidly spreading[42]. One has also keep in mind that the current outbreak is the largest EBOV outbreak thus far with cases rapidly increasing to over 22,000, and thus reliable comparison with previous outbreaks that involved only 1-500 cases, is impossible without any further extrapolation. Lastly, patients treated in

hospitals are likelier to survive than those not treated. This may be just another contributing factor that is likely to misleadingly deflate the CFR.

Conclusion

While we know that GP inflicts vascular damages by its cytotoxicity and that the various EBOV species have different levels of virulence, an understanding of the real cause of virulence especially among the EBOV types remains elusive. The research presented in this paper addresses this enigma and shows that there is a strong correlation between the nucleocapsid disorder and EBOV virulence. This finding sheds lights on the possible reasons and mechanisms of EBOV virulence without discounting what is experimentally known. High intrinsic disorder has been known to provide the means for the promiscuous protein-protein interactions and the efficiency of the transcription process. Therefore, it is likely that disorder is used by the minor and major nucleocasid proteins of the Ebola virus in their functions related to transcription, viral particle budding, and transport of viral proteins. The efficient fulfillment of these roles is likely a key to producing more virus copies and, therefore, guarantees greater quantities of GP from a larger variety of cell-types including those found in vital organs. A better understanding of EBOV virulence would certainly be invaluable in the planning and implementation of measures to control the current epidemic. Besides, a more comprehensive knowledge of the molecular mechanisms of EBOV virulence would also lead to new and more effective strategies in vaccine development.

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.

References

1. G. F. Brookes, J. S. Butel and S. A. Morse, *Jawetz, Melnick, & Aldelberg's Medical Microbiology*, 23 edn., McGraw-Hill, 2004.
2. N. Acheson, *Fundamental of molecular virology*, Wiley, 2007.
3. E. Mühlberger, M. Weik, V. Volchkov, H. Klenk and S. Becker, *J Virol.*, 1999, **73**, 2333-2343.
4. CDC, *Ebola Hemorrhagic Fever Information Packet*, Centers for Disease Control and Prevention, 2009.
5. WHO, *WHO experts consultation on Ebola Reston pathogenicity in humans*, World Health Organiztion, Geneva, 2009.
6. WHO, *WHO: Ebola outbreak report*, World Health Organization, 2014.
7. T. R. Frieden, I. Damon, B. P. Bell, T. Kenyon and S. Nichol, *The New England journal of medicine*, 2014, **371**, 1177-1180.
8. WHO, *Marburg haemorrhagic fever*, World Health Organization, 2012.
9. Z. Peng, J. Yan, X. Fan, M. Mizianty, B. Xue, K. Wang, H. G, V. Uversky and L. Kurgan, *Cell Mol Life Sci*, 2014, 8.
10. B. Xue, A. Dunker and V. Uversky, *J Biomol Struct Dyn.*, 2012, **20**, 137-149.
11. V. N. Uversky, J. R. Gillespie and A. L. Fink, *Proteins Struct. Funct. Genet.*, 2000, **41**, 415-427.
12. A. K. L. Dunker, J. D., C. J. Brown, R. M. William, P. Romero and J. S. Oh, *J Mol. Graph Model*, 2001, **19**, 26-59.
13. V. Vacic, V. Uversky, A. Dunker and S. Lonardi, *BMC Bioinformatics*, 2007, **8**, 211.
14. P. Radivojac, L. Iakoucheva, C. Oldfield, Z. Obradovic, V. Uversky and A. Dunker, *Biophys J*, 2007, **92**, 1439-1466.
15. E. Garner, P. Romero, A. K. Dunker, C. Brown and Z. Obradovic, *Genome Informatics*, 1999, **10**, 41-50.
16. X. Li, P. Romero, M. Rani, A. K. Dunker and Z. Obradovic, *Genome Inform Ser Workshop Genome Inform*, 1999, **10**, 30-40.
17. B. Xue, D. Blocquel, J. Habchi, A. V. Uversky, L. Kurgan, V. N. Uversky and S. Longhi, *Chemical reviews*, 2014, **114**, 6880-6911.

18. G. K. M. Goh, A. K. Dunker and V. Uversky, *BMC Genomic*, 2008, **9**, S4.
19. G. K. Goh, A. K. Dunker and V. N. Uversky, *Virol. J*, 2009, **6**, 69.
20. G. Goh, A. Dunker and V. Uversky, *J Pathog*, 2012, **2012**, 738590.
21. G. Goh, A. Dunker and V. Uversky, *PLOS Currents Outbreak*, 2013.
22. V. Uversky, A. Roman, C. Oldfield and A. Dunker, *J Proteome Res*, 2006, **5**, 1829-1842.
23. B. Xue, M. Mizianty, L. Kurgan and V. Uversky, *Cell Mol Life Sci*, 2012, **69**, 1211-1259.
24. B. Xue, K. Ganti, A. Rabionet, L. Banks and V. Uversky, *Curr Pharm Des*, 2014, **20**, 1274-1292.
25. X. Fan, B. Xue, P. Dolan, D. LaCount, L. Kurgan and V. Uversky, *Mol Biosyst*, 2014, **10**, 1345-1363.
26. G. K. M. Goh, V. Uversky and A. K. Dunker, *Virol. J.*, 2008, **5**, 126.
27. B. Rosner, *Fundamental of biostatistics*, Duxbury, Pacific Grove, CA, 200.
28. G. F. Brooks, K. C. Carrol, J. S. Butel and S. A. Morse, *Jawetz, Melnick, & Aldelberg's Medical Microbiology*, 26 edn., McGraw-Hill, 2012.
29. Z. Yang, H. Duckers, N. Sullivan, A. Sanchez, E. Nabel and G. Nabel, *Nat Med.*, 2000, **6**, 886-889.
30. N. Alazard-Dany, V. Volchkova, O. Reynard, C. Carbonnelle, O. Dolnik, M. Ottmann, A. Khromykh and V. Volchkov, *J Gen Virol*, 2006, **87**, 1247-1257.
31. B. Hartlieb, l. T. Muzio, W. Weissenhorn and S. Becker, *Proc Natl Acad Sci U S A.*, 2007, **104**, 624-629.
32. Y. Liu, S. Stone and R. Harty, *J Infect Dis*, 2011, **Suppl 3**, S817-824.
33. T. Noda, H. Ebihara, Y. Muramoto, K. Fujii, A. Takada, H. Sagara, J. Kim, H. Kida, H. Feldmann and Y. Kawaoka, *PLoS Pathogen*, 2006, **2**, e99.
34. T. Noda, S. Watanabe, H. Sagara and Y. Kawaoka, *J Virol.*, 2007, **81**, 3554-3562.
35. B. Xue, R. Williams, C. Oldfield, G. Goh, A. Dunker and V. Uversky, *Protein Pept Lett.*, 2010, **17**, 932-951.
36. H. J. Dyson and P. E. Wright, *Nat Rev Mol Cell.Biol*, 2005, **3**, 197-2005.
37. P. E. Wright and H. J. Dyson, *J Mol Biol*, 1999, **293**, 321-331.
38. P. Dziubańska, U. Derewenda, J. Ellena, D. Engel and Z. Derewenda, *Acta Crystallogr D Biol Crystallogr*, 2014, **70**, 2420-2429.
39. J. Liu, N. Perumal, C. Oldfield, E. Su, V. Uversky and A. Dunker, *Biochemistry*, 2006, **45**, 6873-6888.
40. S. Baize, D. Pannetier, L. Oestereich, T. Rieger, L. Koivogui, N. Magassouba and e. al, *N Engl J Med.*, 2014.
41. A. Kucharski and W. Edmunds, *Lancet*, 2014, **384**, 1260.
42. S. Gire, A. Goba, K. Andersen, R. Sealfon, D. Par and et al, *Scinece*, 2014, **345**, 1368-1372.

Figure legends

Figure 1. A comparison of PIDs by protein shell levels and virus with references to the average mortality rates by virus types. Disorder levels of minor and major capsid. (Regression Analysis: ($p < 0.01$, $F= 40.2$, $r^2= 0.925$. Average case-fatality rates across EBOV species are also shown. Only outbreaks with more than 50 cases are used in the computation (Oneway ANOVA: $F= 49$, $p<0.01$).

Figure 2. Three dimensional structural representation of portions of the Ebola minor nucleoccapsid (VP30) with red color regions corresponding to fragments predicted as disordered by PONDR® VLXT. A. The VP30 of the ZEBOV. B. REBOV.

Figure 3. The PONDR® VLXT plots of the nucleocapsid proteins NP (A, B) and VP30 proteins (C, D) of two types of Ebola virus, non-pathogenic REBOV and highly virulent. Residues with PONDR® VLXT scores 0.5 and above are considered disordered.

Table 1. A comparison of disorder levels of the minor (VP30) and major (NP) nucleocapsid proteins across EBOV types with CFR.

Virus	Outbreaks/Strains	Case Fatality Rate (CFR, %)[a]	PID (%) NP	PID (%) VP30	UniProt accession numbers
REBOV	Reston 1989 Philippines 1996	Non-pathogenic	38.1±0	33.2±0[c]	Q8JPY1 (1989); Q91DE1 (1996) Q8JPX6 (1989); Q91DD6 (1996)
BEBOV	Bundibugyo, Uganda 2007	25.2±0	40.0±0	33.3±0	B8XCM7 B8XCN3
SEBOV	Sudan/Boniface 1976 Uganda 2000	53.0±0	42.2±0.8	33±0	Q9QP77 (1976); Q5XX08 (2000) B0LPL9 (1976); Q5XX03 (2000)
ZEBOV	Zaire/Mayinga 1976 Gabon/Kikwit 1994-95 Guinea 2014	78.5±9.7	42±1	41.0±0	P18272 (1976); Q9QCE9 (1994); X5GXS8 (2014) Q05323 (1976); Q77DJ5 (1994); A9QPM2 (2014)
Marburg	Popp 1967 Kenya/Ravn 1987	83.2±3.5	46±2	32±1	P35263 (1967); Q1PDD0 (1987) P41326 (1967); Q1PDC6 (1987)

[a] CFR are based on the average CFR among the outbreaks that involve sufficiently large sample size (>50). Further information on the calculations of CFR can be found at: http://www.cdc.gov/ncidod/dvrd/spb/mnpages/dispages/Fact_Sheets/Ebola_Fact_Booklet.pdf

[b] Standard errors are denoted by "±". Percentage of Intrinsic Disorder (PID).

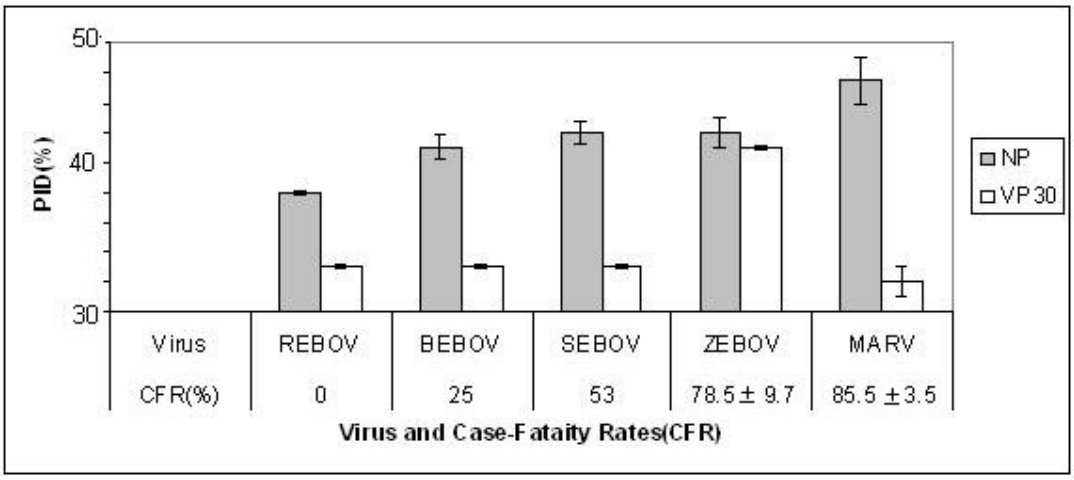
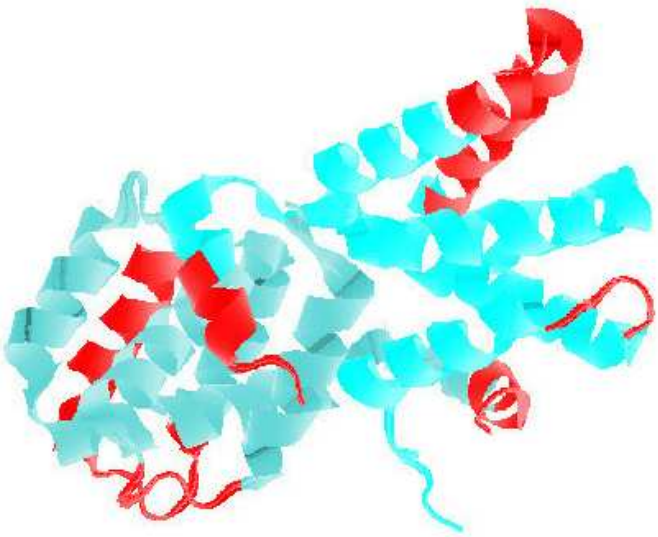
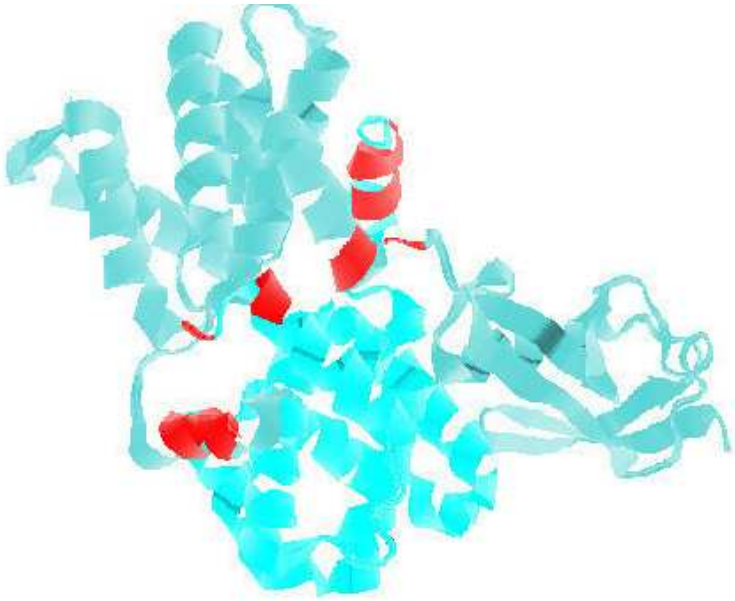


Figure 1

.



Zaire Ebola, VP30 (Minor Nucleocapsid),
PDB: 2i8b, PID: 41±0.1



Reston-Ebola, VP30 (Minor Nucleocapsid),
PDB: 3v7o, PID: 33±0.1

Figure 2

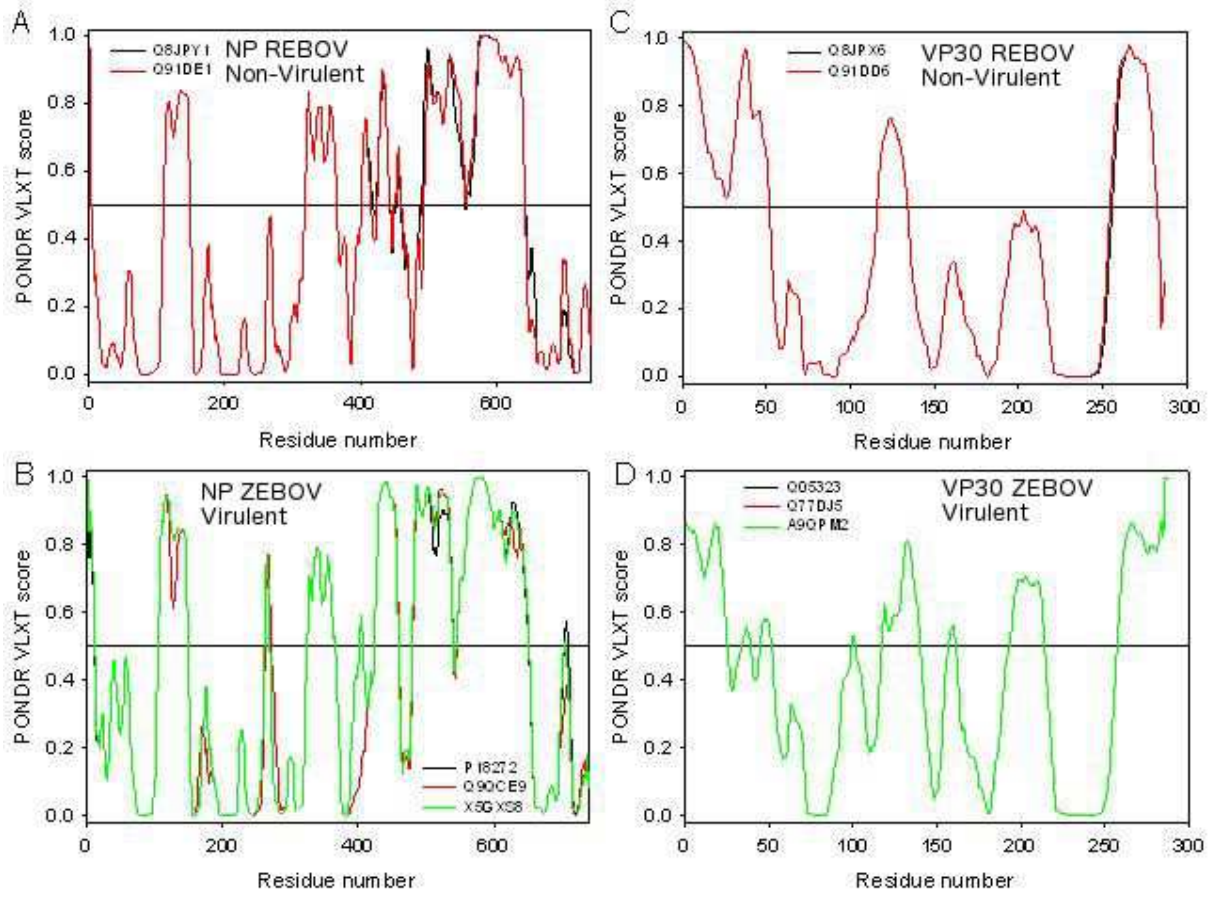


Figure 3