Abstract

Motivation: Ebola viruses are not pathogenic but can be adapted to replicate and cause disease in rodents. Here, we used a structural bioinformatics approach to analyze the mutations associated with Ebola virus adaptation to rodents to elucidate the determinants of host-specific Ebola virus pathogenicity.

Results: We identified 33 different mutations associated with Ebola virus adaptation to rodents in the proteins GP, NP, L, VP24 and VP35. Only VP24, GP and NP were consistently found mutated in rodent-adapted Ebola virus strains. Fewer than five mutations in these genes seem to be required for the adaptation of Ebola viruses to a new species. The role of mutations in GP and NP is not clear. However, three VP24 mutations located in the protein interface with karyopherin α5 may enable VP24 to inhibit karyopherins and subsequently the host interferon response. Three further VP24 mutations change hydrogen bonding or cause conformational changes. Hence, there is evidence that few mutations including crucial mutations in VP24 enable Ebola virus adaptation to new hosts. Since Reston virus, the only non-human pathogenic Ebolavirus species circulates in pigs in Asia, this raises concerns that few mutations may result in novel human pathogenic Ebolaviruses.

Contact: m.n.wass@kent.ac.uk, m.michaelis@kent.ac.uk or j.s.rossman@kent.ac.uk

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

The recent Ebola virus outbreak in West Africa was the first outbreak of a member of the Ebolavirus family that reached epidemic size in humans (de La Vega et al., 2015). It resulted in more than 11 000 reported deaths but this is likely to be an underestimate (Meltzer et al., 2014).

There are currently five known members of the Ebolavirus family. They are: Ebola virus (formerly Zaire virus), Sudan virus, Bundibugyo virus, Tai forest virus and Reston virus. Four of them are pathogenic in humans. The exception is Reston virus, which is not pathogenic in humans. Reston outbreaks in nonhuman primates have occurred in Reston, USA (1989), Sienna, Italy (1992) and Texas (1996) but have not caused disease in humans despite individuals showing IgG antibodies for Reston virus (Miranda and Miranda 2011).

Ebolavirus pathogenesis is caused by a combination of direct cytopathogenic effects and the deregulation of the immune response resulting in massive inflammation (‘cytokine storm’). A number of Ebolavirus proteins have been associated with virus pathogenicity. The viral spike protein GP mediates virus entry into host cells. GP and the matrix protein VP40 have been implicated in causing tissue damage. Shed GP (variants) may contribute to the massive cytokine/chemokine induction. VP35 and VP24 interfere with interferon signaling (Falasca et al., 2015).

Due to the insusceptibility of rodents to Ebola virus infection and disease, rodent models of Ebola virus disease need to be established by virus adaptation via serial passaging (Willyard, 2014). A number of studies have reported on the genetic changes associated with Ebola virus strain adaptation to mice, guinea pigs and hamsters (Cross et al., 2015; Dowall et al., 2014; Ebihara et al., 2006; Volchkov et al., 2000). Here, we applied an in silico approach to
identify structural elements that are most relevant to the development of Ebola virus pathogenicity in a new host.

2 Results

Our analysis focusses on three studies that adapted Ebola viruses to guinea pigs (Cross et al., 2015; Dowall et al., 2014; Volchkov et al., 2000) and one that adapted Ebola virus to mice (Ebihara et al., 2006). Three studies sequenced the virus once it had become pathogenic (Cross et al., 2015; Ebihara et al., 2006; Volchkov et al., 2000). Dowall et al., (2014) sequenced the virus after each passage. Methods are detailed in the Supplementary Material.

2.1 Mutations that are not retained during passaging

24 of the 42 mutations that occurred during the passaging process but reverted to wild type in later passages (Dowall et al., 2014) could be mapped onto protein structures. Many of these mutations are likely to be destabilizing to Ebolavirus proteins (Fig. 1A,B, Supplementary Tables S2 and S3). Mutations that are not retained tend to have lower BLOSUM substitution scores (indicating that such amino acid substitutions do not occur frequently during evolution) than the adaptation mutations (Fig. 1A). Additionally, four of the non-retained mutations are predicted by mCSM (Pires et al., 2014) to be highly destabilizing (ΔΔG < −2.5 kcal/mol) compared to one adaptation mutation (Fig. 1C, Supplementary Tables S2 and S4). All non-retained mutations occur at positions that are completely conserved in the different Ebolavirus species (Fig. 1D, Supplementary Fig. S5). Only two of the 42 (8%) non-retained adaptation-associated changes have been observed in Ebolaviruses, whereas 9 of the 35 (27%) retained changes have been observed in existing Ebolavirus genomes. This further suggests that the non-retained mutations are deleterious, whereas mutations that are retained may have little effect or modulate function as they occur in variable positions.

In NP, both W191R and V323D are predicted to be highly destabilizing to the protein (ΔΔG of −2.973 and −3.339 kcal/mol respectively). Mutation of W191 to arginine would introduce a charged residue in the interior of the protein in a hydrophobic region (Fig. 1E). This may also alter the hydrogen bond that W191 forms with E61, although arginine retains functional groups to hydrogen bond with E61. Similarly, V323D introduces a charged residue in a buried region that is largely hydrophobic apart from H327 and E351, which form a hydrogen bond (Fig. 1F). V323D introduces further negative charge and a hydrogen bond acceptor into this region, which is likely to alter the protein conformation (Fig. 1F).

In VP40, M259R introduces a larger, charged side chain, in a region that is surrounded largely by hydrophobic residues. Arginine at residue 259 could form hydrogen bonds with N257 and potentially other adjacent sidechains. Therefore, M259R may substantially alter VP40 structure.

Temporary changes in VP30, L214P and Q248R, are also likely to affect VP30 structure and function (Fig. 1G). L214 is buried and located in the last turn of an alpha helix. Mutation to proline is likely to shorten the helix resulting in conformational change. Q248R is in the VP30 homodimer interface site (Fig. 1G). The backbone of adjacent residues L247 and L249 form hydrogen bonds with the other subunit (Hartlisch et al., 2007) (Fig. 1G). Although Q248R is a conservative substitution, the increase in charge and size of the amino acid in proximity to the interface may affect VP30 dimer stability.

2.2 Overview of the adaptation-associated mutations

33 unique protein coding mutations were identified in the rodent-adapted Ebola virus genomes. Mutations in GP, NP and VP24 were observed in each of the separate studies (Table 1). Only in some studies, mutations were in L (three studies) or VP35 (two studies) observed. Only mutations that were not retained occurred in the remaining proteins, VP30 and VP40 (Dowall et al., 2014).

Two mutations were observed in multiple studies. The GP I554T mutation was observed in both the Ebihara et al. (2006) and Cross et al., (2015) studies, while VP24 L26F was observed in the Dowall et al., (2014) and Cross et al., (2015) studies. Reverse genetics studies have associated VP24 L26F with increased virulence in rodents (Mateo et al., 2011a,b). 22 of the 33 unique mutations were mapped onto protein structures or models. While there is a structure or model available for all of the Ebola virus proteins, for some of them...
Table 1. Mutations identified in Ebola virus rodent adaptation experiments

<table>
<thead>
<tr>
<th>Protein</th>
<th>Ebihara</th>
<th>Dowall</th>
<th>Volchkov-1</th>
<th>Volchkov-2</th>
<th>Cross</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>N/A</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>GP</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>N/A</td>
<td>1</td>
<td>7†</td>
</tr>
<tr>
<td>L</td>
<td>1</td>
<td>11</td>
<td>1</td>
<td>N/A</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>VP24</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>8§</td>
</tr>
<tr>
<td>VP30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VP35</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>N/A</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>VP40</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>16</td>
<td>6</td>
<td>15</td>
<td>5</td>
<td>35</td>
</tr>
</tbody>
</table>

*VP24 L26F and GP I554T were present in two studies so total number of unique mutations is 7 for VP24 and 6 for GP. Two different adaptation experiments were performed in Volchkov et al., and these are listed separately in the table. §data is only available for VP24 mutations in Volchkov-2.

2.3 Effects of mutations in the glycoprotein (GP)

GP mediates host cell entry (Groseth et al., 2012). GP binding to the endosomal membrane protein NPC1 appears to be required for membrane fusion, but the exact mechanisms remain only partially understood (Miller et al., 2012). Moreover, GP is heavily glycosylated (Lennemann et al., 2014, 2015), which further aggravates the interpretation of the functional consequences of mutations. Across the four studies six different mutations were observed in GP (Table 1). The number of mutations per study range from one (Cross – I554T and Volchkov) to three (Ebihara et al.), while two mutations were identified in Dowall study (Table S4).

Four of the six mutations were mapped onto the GP protein structure (Supplementary Fig. S1, Tables S4 and S5). There is no clustering of the mutations, with each of them being present in a different region of the protein. S65 is a buried residue in the N terminal receptor binding domain. S65P introduces proline into the middle of a beta sheet. This is likely to alter/disrupt the beta sheet and will result in the loss of a hydrogen bond with E100 (Supplementary Fig. S1) resulting in a conformational change within GP. D49N is located at the edge of the interface between GP1 and GP2 (Supplementary Fig. S1). The D49 side chain is not present in the crystal structure suggesting that it is flexible. It could form a hydrogen bond with NS95. The D49N mutation would reduce the charge, but asparagine would still enable the hydrogen bond to be formed. S246P is located on a surface loop near the area that binds the host cell membrane. I544T was observed in two independent studies (Table S4). It is not clear how these changes would modulate GP function. 246 and 544 are variable positions. P246 is conserved in Bundibugyo viruses, T544 in Reston viruses, (Supplementary Fig. S4). Therefore, changes at these positions may be tolerable. S65 and D49, however, are conserved among all Ebolavirus species (Supplementary Fig. S4). Hence, changes in these conserved regions may contribute to Ebola virus adaptation to rodents.

2.4 Effects of mutations in the nucleoprotein (NP)

NP binds to VP30 and VP35 to form with L the viral replication machinery (Groseth et al., 2012; Hartlieb et al., 2007). NP also inhibits the cellular detection of viral RNA and in turn the cellular innate immune response and the destruction of viral RNA by ribonucleases (Fields et al., 2013). In three of the four studies a single mutation was present in NP, while two mutations were identified in the Dowall study. Three of the five mutations in NP could be mapped onto the protein structure (Supplementary Fig. S2, Tables S4 and S6). Adjacent residues S647 and F648 (mutations: S647Y and F648L) in the C terminal domain were mutated in separate studies and the effect of these mutations, if any was not clear (See Supplementary Material). Additionally, leucine is present at residue 648 in Sudan and Reston viruses and position 647 is variable among the Ebolavirus species (Supplementary Fig. S4) indicating that these positions tolerate changes. Mutation S72G is in a highly conserved position and S72 forms a hydrogen bond with the backbone of P42, which is lost on mutation to glycine (Supplementary Fig. S2) but the consequence of this is not obvious. Hence, clear evidence that the NP mutations may be crucial for Ebola virus adaptation to rodents is lacking.

2.5 Multiple mutations in VP24 are likely to be associated with ebola virus pathogenicity

VP24 is involved in the formation of the viral nucleocapsid, the regulation of virus replication, and prevention of interferon signaling (Feldmann and Geisbert, 2011; Mateo et al., 2011a,b; Reid et al., 2006; Watt et al., 2014). VP24 binds to STAT1 and the karyopherins α1 (KPNA1), α5 (KPNA5) and α6 (KPNA6) (Xu et al., 2014). This binding prevents nuclear accumulation of phosphorylated STAT1 and inhibits interferon signaling.

The number of mutations present in VP24 for each separate study ranges from one to three, with one mutation identified in Dowall, Ebihara and second the second Volchkov experiment, two by Cross and three by Volchkov (Table S4). Structural analysis using the complex of VP24 with human KPNA5 provided insight into the likelihood effects of six of the seven VP24 mutations found in rodent-adapted Ebola virus strains (Table S8). Three mutated residues (H186Y, T187I, K142E) are present in or adjacent to the interface site with human KPNA5 (Fig. 2). These mutations were identified in three independent experiments (two in Volchkov et al. and one in Cross et al., Table S4). Wild type H186 forms a hydrogen bond in the interface with residue T434 in human KPNA5 (Fig. 2B). The hydroxyl group in the mutated tyrosine would still be able to form a hydrogen bond with KPNA5 T434 and may enable its interaction with rodent karyopherins. Tyrosine is present at position 186 in Sudan viruses (Supplementary Fig. S4) suggesting that this change is not deleterious. T187I removes intramolecular hydrogen bonds with the backbone of residues H186 and E203 (Fig. 2C). This is likely to increase flexibility in this area, which may allow for binding to additional or different targets. K142E is adjacent to the human KPNA5 interface site and mutations in K142 have been shown to modulate effects on interferon signaling (Ilinsky et al., 2015). This mutation reverses the charge of the side chain and could result in local conformational changes. Moreover, K142 is conserved among all Ebolavirus species (Supplementary Fig. S4), which suggests that changes at this position are likely to have functional consequences. Hence, it is likely that these mutations enable or alter the interaction of VP24 with rodent karyopherins.

The other three mutations (L26F, T50I and L147P) all have an effect on the structure of VP24. Both L26F and L147P are close in space to K142E (Fig. 2A), L26F was observed in two studies (Dowall and Cross) while K142E was observed by Volchkov et al. mCSM predicted L26F to have the most destabilizing effect on
VP24 (Table S4). L26 is located at the end of an alpha helix and is packed against two other alpha helices, resulting in the side chain being largely buried (Fig. 2D). Given the tight packing, mutation to a larger sidechain requires conformational change to accommodate the increased size, although there is no indication of how this would affect VP24 function. However, given that this mutation is conserved among the *Ebolavirus* species, was observed in two independent adaptation experiments (Cross et al., 2015; Dowall et al., 2014) and also in reverse genetics studies (Mateo et al., 2011a,b), it is likely that it has a role in the adaptation to rodent hosts.

T50I and L147P both occur at positions that vary. T50I removes intramolecular hydrogen bonds with the VP24 backbone residues Q36 and K52 (Fig. 2E), which is likely to increase flexibility and, hence, the potential for interaction with additional binding partners. L147P is located towards the end of an alpha helix. Mutation to proline is likely to break the helix, leading to conformational change. Taking these factors into account, the evidence from our structural analysis, the pre-existing knowledge about the VP24 structure and function, and the consistency of VP24 mutations in rodent-adapted Ebola virus strains suggest that VP24 mutations are necessary to facilitate the interaction with STAT1 and/or the karyopherins and in turn inhibition of interferon signaling in a novel host.

Without the complex of VP24 bound to a guinea pig or mouse karyopherin we used mCSM to predict the effect of each of the adaptation mutations on the affinity of VP24 with human KPNA5. The mutations near the interface (K142E, L147P, H186Y and T187I) are predicted to have a destabilizing effect of between 1 and 1.5 kcal/mol on the affinity of the complex, while the other mutations are all predicted to have a destabilizing effect on the complex but with a smaller ΔΔG (Table 2). These results suggest that the adaptation mutations, particularly those close to the interface, are likely to have an effect on the binding of VP24 to rodent karyopherins, however the proposed effect would be to increase the affinity rather than reducing it.

### 3 Discussion

Our analysis suggests that only very few mutations are necessary for Ebola viruses to become pathogenic in a novel host. The different rodent adaptation experiments resulted in 5, 6, 7 or 16 mutations (Table 1 and Supplementary Table S4). Only mutations in VP24, GP and NP were observed in all adaptation studies. Per experiment, a maximum of 5 mutations were detected across these three genes, representing an upper bound on the minimum number of coding mutations that enable Ebolaviruses to cause disease in a novel, previously non-susceptible host.

While GP and NP are mutated in each study and may have a role in adaptation, there is no clear evidence that mutations in them are critical for Ebola virus adaptation to a new host. GP and NP have (together with L) the greatest variability in their sequences (Jun et al., 2015) and some of the mutations are in variable positions. Therefore, variation in these sequences may not be surprising and may not necessarily reflect Ebola virus adaptation to a new host.

In contrast to GP and NP, there is strong evidence that mutation of VP24 is vital for Ebola virus adaptation to a novel host. The VP24 mutations H186Y, T187I and K142E occur in a region that is critical for karyopherin binding and the modulation of the host cell interferon response. Hence, these changes seem to be required to...
inhibit the interferon response in rodents. This agrees with evidence suggesting that VP24 may be a determinant of species-specific pathogenicity between different Ebolaviruses (Zhang et al., 2012). Conserved differences in VP24 residues located in the VP24-karyopherin interface site (similarly to the adaptation mutations) that differ between Reston viruses, the only Ebolaviruses that are not pathogenic in humans, and the other Ebolaviruses have been proposed to critically contribute to the lack of human pathogenicity of Reston viruses (Pappalardo et al., 2016).

In conclusion fewer than five mutations are required for Ebolavirus adaptation to a new species. The potential for Ebolaviruses to adapt to novel host species is of relevance for two reasons. Firstly, circulation in (domestic) species may increase the exposure of humans to Ebolaviruses. Reston viruses circulate in pigs and can be transmitted from pigs to humans (possibly by air) (Atherstone et al., 2015; Barrette et al., 2009; Marsh et al., 2011; Miranda and Miranda, 2011; Olson et al., 2012; Osterholm et al., 2015; Pan et al., 2014; Weingartl et al., 2013). Hence, human pathogenic Reston viruses may emerge. Secondly, Ebolavirus adaptation to new species (including domestic species like pigs and dogs (Atherstone et al., 2015; Barrette et al., 2009; Marsh et al., 2011; Miranda and Miranda, 2011; Olson et al., 2012; Osterholm et al., 2015; Pan et al., 2014; Weingartl et al., 2013) may result in greater virus diversity and, in turn, result in an increased number of Ebolaviruses that pose a potential threat to humans.

Conflict of Interest: none declared.

References


Fields, B.N. et al. (2013) Fields Virology Lippincott Williams & Wilkins.


