Characterizing alpha helical properties of Ebola viral proteins as potential targets for inhibition of alpha-helix mediated protein-protein interactions [version 1; referees: 1 approved with reservations]

Sandeep Chakraborty¹,², Basuthkar J. Rao², Bjarni Asgeirsson³, Abhaya M. Dandekar¹

¹Plant Sciences Department, University of California, Davis, CA, 95616, USA
²Department of Biological Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai, 400 005, India
³Science Institute, Department of Biochemistry, University of Iceland, Dunhaga 3, IS-107 Reykjavik, Iceland

Abstract

Ebola, considered till recently as a rare and endemic disease, has dramatically transformed into a potentially global humanitarian crisis. The genome of Ebola, a member of the Filoviridae family, encodes seven proteins. Based on the recently implemented software (PAGAL) for analyzing the hydrophobicity and amphipathicity properties of alpha helices (AH) in proteins, we characterize the helices in the Ebola proteome. We demonstrate that AHs with characteristically unique features are involved in critical interactions with the host proteins. For example, the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain has an AH with a large hydrophobic moment. The ability of this AH to bind to other host proteins is disrupted by a neutralizing antibody derived from a human survivor of the 1995 Kikwit outbreak, emphasizing the critical nature of this secondary structure in the virulence of the Ebola virus. Our method ensures a comprehensive list of such 'hotspots'. These helices probably are or can be the target of molecules designed to inhibit AH mediated protein-protein interactions. Further, by comparing the AHs in proteins of the related Marburg viruses, we are able to elicit subtle changes in the proteins that might render them ineffective to previously successful drugs. Such differences are difficult to identify by a simple sequence or structural alignment. Thus, analyzing AHs in the small Ebola proteome can aid rational design aimed at countering the 'largest Ebola epidemic, affecting multiple countries in West Africa' (http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/index.html).
This article is included in the Disease Outbreaks gateway.

Associated Research Note


Corresponding author: Sandeep Chakraborty (sanchak@ucdavis.edu)

Competing interests: No competing interests were disclosed.

Grant information: AMD wishes to acknowledge grant support from the California Department of Food and Agriculture PD/GWSS Board. BJ acknowledges financial support from Tata Institute of Fundamental Research (Department of Atomic Energy). Additionally, BJR is thankful to the Department of Science and Technology for the JC Bose Award Grant. BA acknowledges financial support from the Science Institute of the University of Iceland.

Copyright: © 2014 Chakraborty S et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

How to cite this article: Chakraborty S, Rao BJ, Asgeirsson B and Dandekar AM. Characterizing alpha helical properties of Ebola viral proteins as potential targets for inhibition of alpha-helix mediated protein-protein interactions [version 1; referees: 1 approved with reservations] F1000Research 2014, 3:251 (https://doi.org/10.12688/f1000research.5573.1)

Introduction

The Ebola virus was first discovered in 1976¹, and has been since known as a rare, but deadly disease². However, the current outbreak in West African countries (Guinea, Liberia, Nigeria, Sierra Leone and Senegal) has rapidly deteriorated into a full blown epidemic³, and poses grave humanitarian dangers to these countries⁴. Ebola, along with the Marburg virus, belongs to the Filoviridae family⁵, and causes haemorrhagic fever⁶ by quickly suppressing innate anti-viral immune responses to facilitate uncontrolled viral replication⁷.

Interestingly, the genome of the Ebola virus encodes seven proteins⁸, although their extreme ‘plasticity allows multiple functions’⁹,¹⁰. Protein structures are formed by well ordered local segments, of which the most prevalent are alpha helices (AH) and β sheets. AHs are right-handed spiral conformations which have a hydrogen bond between the carbonyl oxygen (C=O) of each residue and the alpha-amino nitrogen (N-H) of the fourth residue away from the N-terminal. AH domains are often the target of peptides designed to inhibit viral infections¹⁰–¹². Recently, we have provided open access to software that has reproduced previously described computational methods¹³ to compute the hydrophobic moment of AHs (PAGAL¹⁴).

In the current work, we characterize the helices in the Ebola proteome using PAGAL, and demonstrate that the helices with characteristically unique feature values are involved in critical interactions with the host proteins. The PDB database is queried for the keyword ‘Ebola’, and the structures obtained are analyzed using DSSP (Define Secondary Structure of Proteins)¹⁵ for identifying AHs. We process all PDB structures, and do not filter out redundant structures based on sequence. These helices are analyzed using PAGAL, and the results are sorted based on three criteria - hydrophobic moment and high proportion of positive or negative residues. The helices that are ranked highest in these sorting criteria are involved in critical interactions with either antibodies or host proteins. For example, the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain has an AH with the largest hydrophobic moment in all helices analyzed¹⁶. This helix is disrupted by a neutralizing antibody derived from a human survivor of the 1995 Kikwit outbreak, emphasizing the critical nature of this helix in the virulence of Ebola¹⁷. Another example, obtained by choosing the helix with the highest proportion of negatively charged residues, is the interaction between the human karyopherin alpha nuclear transporters C terminus and the Ebola virus VP24 protein (eVP24)¹⁸, which suppresses tyrosine-phosphorylated STAT1 nuclear import¹⁹. These helices probably are, or can be, the target of molecules designed to inhibit AH mediated protein-protein interactions²⁰. Our method provides a comprehensive list of such targets. Further, each protein can be individually queried using PAGAL, and thus identified helices that might have a poor global rank, but still be critical in the particular proteins context.

Although, Ebola and Marburg viruses are members of the Filoviridae family²¹, they have different antigenicity of the virion glycoprotein²². These differences are probably the rationale for the lesser mortality observed in Marburg outbreaks. By comparing the AHs in proteins of Marburg and Ebola viruses, we are able to elicit subtle changes in the proteins that might render them ineffective against previously successful drugs. These differences are not apparent from a simple sequence or structural alignment. Thus, in the current work, we elucidate a simple methodology that can aid rational design of drugs and vaccine, an important aspect of the global effort to counter the deadly Ebola epidemic.

Materials and methods

We searched for the keyword ‘Ebola’ in the PDB database (Table 1). Subsequently, each protein was split based on the chain ID, resulting in 146 single chained proteins (See ALPHA.zip in Dataset 1). We have not reduced the set based on sequence similarity since the proteins might have different conformations based on their ligands. Note, this list might include non-Ebola proteins which might have been co-crystallized with the Ebola protein. However, they have been put through the same analysis since they might provide insights into the Ebola proteins themselves.

These proteins were then analyzed using DSSP²³, and resulted in 758 helices in all (See ALPHA.zip in Dataset 1). These helices were then analyzed using PAGAL. The PAGAL algorithm has been detailed previously²⁴. Briefly, the Edmundson wheel is computed by considering a wheel with centre (0,0), radius 5, first residue coordinate (0,5) and advancing each subsequent residue by 100 degrees on the circle, as 3.6 turns of the helix makes one full circle. We compute the hydrophobic moment by connecting the center to the coordinate of the residue and give it a magnitude obtained from the hydrophobic scale (in our case, this scale is obtained from²⁴). These vectors are then added to obtain the final hydrophobic moment.

The color coding is as follows: all hydrophobic residues are colored red, while hydrophilic residues are colored in blue: dark blue for positively charged residues, medium blue for negatively charged residues and light blue for amides.

The raw file generated by analyzing all 146 proteins through PAGAL is provided as PAGALRAW-DATA.txt (Dataset 1), and contains the hydrophobic moment, percent of positive charges and the total number of charged residues for every helix. These are then sorted based on the charge (negative or positive) or the hydrophobic moment. We ignore the helices that have none or a single one charged residue.

All protein structures were rendered by PyMol (http://www.pymol.org/). The sequence alignment was done using ClustalW²⁵. The alignment images were generated using Seaview²⁶. Protein structures have been superimposed using MUSTANG²⁷.

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1ESP,1H2D,3TCQ,4LDM...</td>
<td>Ebola virus envelope protein Minor nucleoprotein VP30 Polymerase cofactor VP35 Membrane-associated protein VP24 Nucleoprotein Matrix protein VP40</td>
</tr>
</tbody>
</table>

Table 1. PDB ID of Ebola proteins analyzed.
Results and discussion

Dataset 1. PAGAL analysis of Ebola-related alpha helices

http://dx.doi.org/10.5256/f1000research.5573.d37453

A PDB database search using the keyword 'Ebola' generated 146 single chained proteins, which were analyzed using Define Secondary Structure of Proteins, resulting in 758 alpha helices (ALPHA.zip). Note, this list might include non-Ebola proteins which might have been co-crystallized with the Ebola protein. These helices were analyzed using PAGAL (PAGALRAWDATA.txt), which details the hydrophobic moment, percent of positive charges and the total number of charged residues for every helix.

Helices with large hydrophobic moment

We began by analyzing the helices which have a large hydrophobic moment (hydrophobic scale is obtained from\textsuperscript{13}) (Table 2). The Edmundson wheel for the helix 1EBOE.HELIX1 from the structure of GP2 from the Ebola virus membrane fusion glycoprotein (PDBid:1EBO)\textsuperscript{14} is shown in Figure 1a. Figure 1b shows the residues comprising these helices (in magenta) in the apo form (PDBid:1EBO)\textsuperscript{14}. This helix is disrupted by a neutralizing antibody derived from a human survivor of the 1995 Kikwit outbreak (PDBid:3CSY)\textsuperscript{15}, emphasizing the critical nature of this helix in the virulence of Ebola (Figure 1c,d). Table 3 shows the residues in the

Table 2. Identifying helices with unique properties. Property based on which the sorting is done is either the Hydrophobic moment (HM) and the percentage of negative (NEG) or positive residues (POS). HM: Hydrophobic moment, RPNR: Ratio of the positive to the negative residues, Len: length of the helix, NCH: number of charged residues, GP: glycoprotein from Ebola, VP24: Membrane-associated protein from Ebola, VP35: Polymerase cofactor.

<table>
<thead>
<tr>
<th>Property</th>
<th>Protein</th>
<th>Helix</th>
<th>Len</th>
<th>HM</th>
<th>RPNR</th>
<th>NCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>HM</td>
<td>GP</td>
<td>1EBOE.HELIX1</td>
<td>46</td>
<td>16.2</td>
<td>0.5</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1EBOE.HELIX0</td>
<td>29</td>
<td>11.5</td>
<td>0.5</td>
<td>13</td>
</tr>
<tr>
<td>NEG</td>
<td>VP24</td>
<td>4U2XA.HELIX5</td>
<td>16</td>
<td>4.4</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3FKEA.HELIX2</td>
<td>14</td>
<td>3.2</td>
<td>0.2</td>
<td>4</td>
</tr>
<tr>
<td>POS</td>
<td>VP24</td>
<td>4U2XA.HELIX7</td>
<td>19</td>
<td>6.5</td>
<td>0.8</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3FKEA.HELIX1</td>
<td>15</td>
<td>7.8</td>
<td>0.8</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 3. Interactions obtained from the crystal structure of the Ebola virus glycoprotein in complex with a neutralizing antibody from a human survivor. The helix with a large hydrophobic moment, as determined from PDBid:1EBOE, is disrupted in the structure from PDBid:3CSY through possible hydrogen bonds with different residues in the human Fab KZ52 heavy chain (antibody, chain A). The helix residues are: 553-597 in chain J, PDBid:3CSY.

<table>
<thead>
<tr>
<th>AtomEbola</th>
<th>AtomAntibody</th>
<th>Dist (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASP/552/OD1</td>
<td>SER/53/OG</td>
<td>2.5</td>
</tr>
<tr>
<td>GLY/553/N</td>
<td>THR/100/OG1</td>
<td>2.7</td>
</tr>
<tr>
<td>ASN/550/O</td>
<td>ASN/31/O</td>
<td>2.9</td>
</tr>
<tr>
<td>ASP/552/OD1</td>
<td>SER/53/CB</td>
<td>2.9</td>
</tr>
<tr>
<td>ASN/550/ND2</td>
<td>PRO/97/O</td>
<td>3.0</td>
</tr>
<tr>
<td>ASN/550/ND2</td>
<td>ASN/31/O</td>
<td>3.2</td>
</tr>
<tr>
<td>ASP/552/OD2</td>
<td>SER/53/OG</td>
<td>3.2</td>
</tr>
</tbody>
</table>

Figure 1. Helix with large hydrophobic moment in GP2 from the Ebola virus membrane fusion glycoprotein. (a) Edmundson wheel for 1EBOE.HELIX1. The hydrophobic moment vector is not to scale. The color coding is as follows: all hydrophobic residues are colored red, while hydrophilic residues are colored in blue: dark blue for positively charged residues, medium blue for negatively charged residues and light blue for amides. (b) Structure of PDBid:1EBOE, 1EBOE.HELIX1 is marked in magenta and the leucine zipper is in blue. (c) 1EBOE.HELIX1 is disrupted by an antibody derived from a human survivor of the 1995 Kikwit outbreak (PDBid:3CSY). (d) Gly553/N on 1EBOE.HELIX1 makes a possible hydrogen bond to Thr100/OG1 at a distance of 2.7 Å.
specified helix (residues 553–597, chain J, PDBid:3CSY) making possible hydrogen bonds with different residues in the human Fab KZ52 heavy chain (residues 1–228, chain A, PDBid:3CSY). Among all the interactions, only Gly553 is on 1EBOE.HELIX1 (at a distance of 2.7 Å from Thr100/OG1), although the others are sequentially proximal. These few interactions are sufficient to disrupt this helix, rendering the virus non-virulent, and leading to human recovery. The importance of interfacial hydrophobicity in viral proteins involved in host entry through membrane fusion has recently been discussed in details, and remains ‘an underutilized therapeutic target’ \(^{26}\). It is also interesting that the helix is also involved in a disulphide bond after its disruption (Cys556 and Cys511). 1EBOE.HELIX0 (Table 2) also has a high hydrophobic moment, but is actually an isoleucine zipper derived from GCN4 \(^{27}\) (Figure 1b).

**Helices with high proportion of negatively charged residues. Identifying difference among different species**

We then analyzed the helices having a high proportion of negatively charged residues, sorted based on the length of the helix when the percentage of negatively residues are the same (Table 2). Figure 2a shows the Edmundson wheel for the helix 4U2XA.HELIX5 (which has only two charged residues - the basic E113 and D124), while Figure 2b,c shows this helix in the protein complex marked in magenta. Protein PDBid:4U2XD is the human karyopherin alpha nuclear transporter (KPNA) C terminus in complex with the Ebola virus VP24 protein (eVP24)\(^{18}\). eVP24 interferes with the immune response by selectively targeting tyrosine-phosphorylated STAT1 nuclear import\(^{16}\). It does not hinder the transport of other cargo that may be required for viral replication. 4U2XA.HELIX5 is responsible for forming the complex with the KPNA protein through a helix (4U2XD.HELIX9, in blue), and K481 from KPNA is in contact with D124 from eVP24 (distance between K481/NZ and D124/OD2 is 3.98 Å). Their interaction is probably electrostatic, since the atoms have opposite charges. VP24 has also been shown to directly bind to STAT1, further compromising the immune response\(^{28}\).

The next helix having a high proportion of negatively charged residues (3FKEA.HELIX2) is from a VP35, a classic example of a moonlighting protein, that can be a component of the viral RNA polymerase complex, a viral assembly factor, or inhibitor of host interferon production\(^{29}\). We have not been able to identify a critical role for this helix in the protein from current literature. However,
These differences are definitely encoded in the proteins expressed by these viruses, and the design of drugs and vaccines to counter them should take these differences into account.

Helices with high proportion of positively charged residues

For helices having a high proportion of positively charged residues, we could not find any reference to the critical nature of the first helix (Table 2, 4U2XA.HELIX7). This helix is marked in yellow in Figure 2c. The second helix (3FKEA.HELIX1) is from VP35, which was discussed previously\(^29\). This helix spans residues 238–252 and includes Lys248 and Lys251, a basic patch which is ‘100% identical among members of the Ebola viral isolates’\(^29\), and Ala238, Gln241, Leu242, Val245, Ile246, Leu249 which interacts with a \(\beta\) sheet to create a hydrophobic subdomain\(^29\). This helix is marked in magenta in Figure 3b, and the Edmundson wheel is shown in Figure 3c. Once again, we demonstrate that unique values of an AH is a strong indicator of its significance in the viral functionality.

Multifunctional/moonlighting

The multifunctional roles played by many of these Ebola proteins is probably due to stretches of intrinsically disordered regions within...
The Zaire Ebola virus nucleoprotein. It can be seen that 4QAZA.HELIX0 (residues 646–658) has a reasonably high hydrophobic moment (although it will not rank highly if we analyze all helices from the proteome), and also a high number of charged residues (Figure 4a,b). It has been observed that ‘the side chains of Glu645, His646, Glu649, Lys684, Glu695, Glu709, Lys728 and Gln739 are partly disordered so that some or all of their atoms are not visible in the electron density’.

Glue645, His646, Glu649 are part of this helix, and are thus critical to the disorderedness of the protein, which is critical for its moonlighting roles. Note, that Glu has been observed to be the second most disorder promoting residue (after proline).

The above examples have analyzed all helices from the Ebola proteome. However, it also possible to analyze the helices in a single protein, and probe those for unique features. Table 5 shows the values obtained from PAGAL for helices of the C-terminal domain of the Zaire Ebola virus nucleoprotein. It can be seen that 4QZA.HELIX0 (residues 646–658) has a reasonably large hydrophobic moment, and has been hypothesized to be part of the protein which is involved in protein-protein interactions. Further, these helices have residues with disordered sidechains, which are known to be critical for moonlighting functions. HM: Hydrophobic moment, RPNR: Ratio of the positive to the negative residues, Len: length of the helix, NCH: number of charged residues.

Conclusions
The ability of a genome as small as the Ebola virus to inflict a dishearteningly high percentage of mortality in human subjects is a humbling experience in the context of the tremendous technological
advancements achieved in the last few decades\textsuperscript{3,4}. The Ebola virus potently suppresses the human immune response\textsuperscript{24,85} by binding with key human proteins involved in the immune pathway\textsuperscript{15}. These protein-protein interactions are often mediated through well-structured secondary regions within the protein structures (alpha helices), and the design of molecules that inhibit these ‘hotspots’\textsuperscript{20,36} has been a well known strategy to develop drugs to counter bacterial and viral infections\textsuperscript{16-12}. For example, synthetic peptides derived from the oligomerization domain of polymerase subunits has been shown to inhibit viral proteins\textsuperscript{17-38}. On the other hand, there might exist other protein domains that might be exploited by non-native viral peptides to obstruct viral functionality. In the current work, we characterize alpha helices in the Ebola virus proteome using a recently implemented open access software (PAGAL)\textsuperscript{13}, thus identifying potential targets for inhibition of the helix mediated interactions. Through several examples, we demonstrate that helices with unique features are involved in interactions with host proteins (either antibodies from survivors, or proteins regulating the immune response). Further, we also provide an alternate way of analyzing differences in related proteins (from the Marburg virus) by focusing on the properties of corresponding helices. As future work, we intend to develop methodologies to design peptides that would target these ‘hotspots’\textsuperscript{39}. It has to be kept in mind that it has been a challenge to design small ligands that disrupt protein-protein interactions, and designers resort to several innovative techniques to overcome thermodynamic instability or proteolytic susceptibility\textsuperscript{24-41}. These helices can essentially be epitopes\textsuperscript{42-45} for developing antibodies against the virus\textsuperscript{46-47}. Interestingly, ZMapp, a cocktail of three antibodies has shown reversion of advanced Ebola symptoms in non-human primates\textsuperscript{48}, and uses only glycoprotein-specific epitope generated antibodies\textsuperscript{49,50}. It is interesting to hypothesize that additions to this cocktail with antibodies derived from other epitopes (for example, 4U2XA.HELIX5 from VP24 that is involved in immune response suppression) could prove more effective. Thus, we provide a comprehensive list of potential targets from the small proteome of the Ebola virus that can directed rational design to quickly innovate therapies.

Data availability
F1000Research: Dataset 1. PAGAL analysis of Ebola-related alpha helices, 10.5256/f1000research.5573.d37453

Author contributions
SC wrote the computer programs. All authors analyzed the data, and contributed equally to the writing and subsequent refinement of the manuscript.

Competing interests
No competing interests were disclosed.

Grant information
AMD wishes to acknowledge grant support from the California Department of Food and Agriculture PD/GWSS Board. BJ acknowledges financial support from Tata Institute of Fundamental Research (Department of Atomic Energy). Additionally, BJR is thankful to the Department of Science and Technology for the JC Bose Award Grant. BA acknowledges financial support from the Science Institute of the University of Iceland.

References


6. Daugerty MD, Malik HS: How a virus blocks a cellular emergency access lane to the nucleus, STAT1 Cell Host Microbe. 2014; 16(2): 150–152. Published Abstract | Publisher Full Text


Page 8 of 11
Open Peer Review

Current Referee Status: 

Version 1

Referee Report 04 November 2014

https://doi.org/10.5256/f1000research.5956.r6506

Winfried Weissenhorn
Unit for Virus Host-Cell Interactions (UVHCI), Grenoble, France

The authors suggest that they can identify alpha helices and predict their propensities to be targeted by small molecules. Their test case is the small Ebola virus genome, where several crystal structures are available.

First they compute the hydrophobic moment of identified helices with their previously published program PAGAL and classify them based on hydrophobicity, positive or negative charges. They conclude that helices with unique feature values are involved in host protein interaction.

Page 4: It is not correct to state that “this helix is disrupted by a neutralizing antibody derived from a human survivor …“. HR1 or helix 1 from Gp2 is split into 4 small helices in the native GP structure and antibody binding prevents its refolding into the post fusion conformation represented by the Gp2 structure. Now one can argue that small molecules could interfere with the formation of the triple stranded coiled coil formed by HR1 in the post fusion structure. This needs to be clarified in the text.

Next they identified a charged helix in Vps24 that interacts with karyopherin. Why was this chosen? Because of the available structure? This helix contains only two charged residues and would not fall under the classification of carrying a high charge!

The third helices described in detail are from Vps35 and the authors identify several helices with carry charges, but no clear targets are discussed.

Page 6: The authors make a connection between the number of acidic residues in a helix from Ebola Vps35 compared to Marburg Vps35 and the frequency of outbreaks, which is a complete over interpretation of their data.

In summary the manuscript describes an interesting approach to identify or validate potential drug targets. However, the authors need to be more cautious in interpreting their results. Without any experimental validation their approach to link helical properties to protein interaction propensities is extremely weak.

Competing Interests: No competing interests were disclosed.
I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com