Medical Hypotheses 144 (2020) 109925

Contents lists available at ScienceDirect

Medical Hypotheses

journal homepage: www.elsevier.com/locate/mehy

The potential of putative zinc-binding motifs of haemagglutinin (HA) protein for categorization and prediction of pathogenicity of H5 subtypes of avian influenza virus

Issa A. Muraina[a,d,*], Clement A. Meseko[b], Folorunso O. Fasina[c]

[a] Biochemistry Division, National Veterinary Research Institute, Vom, Nigeria
[b] Animal Influenza Division & TADs, National Veterinary Research Institute, Vom, Nigeria
[c] FAO-ECTAD, Dar es Salam, Tanzania & Department of Veterinary Tropical Diseases, University of Pretoria, South Africa
[d] Influenza Resource Centre, National Institute for Biological Standards and Control, Hertfordshire, UK

ABSTRACT

In the present study, we used the potential of bioinformatics and computational analysis to predict the existence and biological relevance of zinc finger (ZF) motifs in heamagglutinin (HA) protein of Avian Influenza (AI) virus. Sequence data of Avian Influenza (AI) viruses were retrieved from accessible databases (GenBank, GISAID, IRD) and analyzed for the existence, as well as functional prediction of the putative zinc finger or "zinc-binding" motif(s) of HA protein. It is hypothesized that the ZF motif(s) in HA of AI virus can be used as a "novel" biomarker for categorization of the virus and/or its virulence. As a model for analysis, we used the H5 subtypes of highly pathogenic, non-pathogenic and low pathogenic avian influenza (HPAI, NPAI and LPAI) viruses of H5N1 and H5N2 of avian and human origins. Interestingly, our method of characterization using the zinc-finger agrees with the existing classification in distinguishing between highly pathogenic and non-pathogenic or low pathogenic subtypes. The new method also clearly distinguished between low and non-pathogenic strains of H5N2 and H5N1 which are indistinguishable by the existing method that utilizes the sequence of the polybasic amino acids of the proteolytic cleavage site for pathogenicity. It is hypothesized that zinc through the activities of zinc-binding proteins modulates the virulence property of the viral subtypes. Our observation further revealed that only the HA protein among the eight encoded proteins of influenza viruses contain high numbers of Cys-His residues. It is expected that the information gathered from the analysis of the data will be useful to generate more research hypotheses/designs that will give further insight towards the identification and control of avian influenza virus through the molecular manipulation of zinc finger motifs present in viral HA protein.

Introduction

Zinc is a metallo-ion essential for organismal growth, survival and normal immune function. Its deficiency causes poor immune system modulation and immune suppression, and can increase the severity and duration of viral infection [1]. Pathogenic viruses have certain proteins that are used in viral replication cycle and one of these important proteins is called zinc-finger or zinc-binding protein [2,3]. Therefore stopping or slowing down of viral replication is possible through the deprivation of living organism of zinc either by using an inhibitor such as zinc-finger ejector [4], or by mutation or disruption of the normal sequences of Cysteine-Histidine (CH)-rich zinc-binding motif in the virus [5]. Several publications have demonstrated the existence of zinc-binding domains in many pathogenic viruses such as the M1-protein and the V-protein (V_u) regions of highly pathogenic avian influenza virus and Newcastle disease virus respectively [2,5,6].

Mutation of the classical or consensus Cys-His putative zinc-binding motif (Cys2His2 or CCHH) in M1-protein of lethal influenza virus as well as in VP30-protein of Ebola virus produced mutant viruses with less replicating and less pathogenic abilities [5,7]. This is possible because of the fact that this region of protein in some viruses acts as transcription activation domain for viral DNA synthesis and for zinc mobilization and binding [7,8]. It was also shown that mutation or substitution of Cysteine or Histidine residue of zinc-binding domain with Alanine in human coronavirus 229E (HCov-229E) non-structural protein 13 (nsp13) caused a more deleterious effect [9]. The zinc-binding domain in this virus and other Nidoviruses is involved in viral replication and transcription by modulating the enzymatic activities of the helicase domain and other yet undiscovered mechanisms. Taking together, the importance of viral zinc regulation in the pathogenesis of viral infections cannot be underestimated.

Avian influenza viruses are classified as high or low pathogenic based on severity and mortality in chickens [10]. In nature, avian influenza viruses in birds exist in combinations of subtypes H1 to 16 and N1 to 11 [11]. All subtypes can be found in waterfowls especially in its low pathogenic forms [12]. The birds that are infected with low

* Corresponding author.
E-mail address: issa.muraina@nvri.gov.ng (I.A. Muraina).

https://doi.org/10.1016/j.mehy.2020.109925
Received 2 April 2020; Received in revised form 15 May 2020; Accepted 28 May 2020

pathogenic virus (LPAI) often show no signs of infection and hardly resulted in a disease. LPAI may later revert to virulence through mutations in the antigenic cleavage site of the haemagglutinin (HA) protein. The amino acid sequences at the proteolytic cleavage site of the HA lacks poly (multiple) basic amino acids and correspond to PQRETGL*F and this is typical of low pathogenic avian influenza (LPAI) as well as non-pathogenic avian influenza (NPAI) viruses. Mutations at this site can cause the virus to acquire potential virulence factor and initiate novel infection that results in highly pathogenic disease (HPAI) in domestic birds [13]. One of the mechanism of virulence is caused by the cleavage of the HA protein by ubiquitous enzymes that may be found in all body tissues and fluid, hence the virus gain ability to replicate and produce pathological lesions [14]. In another experiment by Toshihiro and co-workers [15] with series of passages of avian influenza virus in chicken, the avirulent virus became highly pathogenic in chickens, producing 100% mortality rate. Sequences analysis at the HA cleavage site of the original virus revealed a typical avirulence type of sequence 'RETR', which progresses incrementally to a typical virulence type of sequence 'RRKKR' during repeated passages in chickens. Therefore, continuous avian influenza virus surveillance is thus important for monitoring these genetic changes and biological host adaptations using various biomarkers including zinc finger motif in order to better deduce and interpret virus evolution.

In the present analysis of H5 subtypes of avian influenza virus, we have revealed that the HA protein showed some forms of mutation or substitution in the amino acids sequences of the proposed zinc-binding domains and this may be useful in the categorization of the virus into HPAI, LPAI and NPAI as well as differentiating between the non-virulent LPAI and NPAI that are both lacking the virulent HA cleavage polyybasic amino acid sequences. Although none of the mutations occur on the Cys-His (CH)-rich residues or on the consensus CCHH zinc finger, they are found on other amino acid residues within the different zinc finger motifs.

Hypothesis

The amino acids substitution or mutation occurring in the proposed ZF motif(s) in HA protein of H5 lineage of avian influenza virus may be useful in the categorization of the virus into HPAI, LPAI and NPAI respectively as well as differentiating between the non-virulent LPAI and NPAI that are indistinguishable by the existing method as they both lack the virulent HA proteolytic cleavage (polybasic) amino acid sequences.

Evaluation of the hypothesis

Evaluation of the full-length heamagglutinin (HA) protein sequences in subtypes/lineages of avian influenza viruses in Figs. 1 and 2 clearly indicated the presence of putative zinc-binding or zinc-finger (ZF) motifs or domains. The motifs are grouped in to 5 fragments (I-V) separated by molecules of 2 Histidine (H) residues flanking the centrally located ZF motif III, which places ZF motif I & II on one side and ZF motif IV & V on the other side (Fig. 1). All the 5 fragments are made up of cysteine-histidine (CH)-rich residues except the last fragment (motif V) which is made up of only cysteine (C)-rich residues.

Consequence of the hypothesis and discussion

The 5 classes of motifs in different subtypes (i.e. subtypes 1, 2 & 3 of H5N1 and H5N2) of the same H5 lineage showed similarity or full conservation of colour coding with each other by having similar number of amino acid residues for both highly pathogenic and non-pathogenic or low pathogenic viruses. However this is completely different from a different viral lineage H1 and its subtype such as H1N1, an indication that ZF motifs may be used to distinguish influenza viral lineages at this level of analysis.

The high conservation of these "zinc-binding motifs" in similar lineage implies functional importance for both highly pathogenic, low pathogenic and non-pathogenic subtypes. Although in term of total number of amino acids making up each motif, there is no distinguishing feature between the subtypes of similar lineage of either highly pathogenic, non-pathogenic or low pathogenic viruses as they all have equal number of amino acids composition. Since there is full conservation of zinc-binding motifs in them, it is possible that these viruses (irrespective of their pathogenicity) utilize these motifs to colonize the host and establish themselves for survival. Because the first line of action utilized by the virus is the use of the *haemagglutinin* gene (protein) to attach to the host before invasion and causing disease especially by the highly pathogenic subtypes.

Critical analysis of the composition and types of the accompanying amino acid sequences with the Cys-His residues in each motif of subtypes H5N1 & H5N2 showed that these strains of viruses can be grouped or categorized into distinct highly pathogenic (HPAI) and non-pathogenic (NPAI) H5N1 subtypes (Appendix Table 1) or low pathogenic H5N2/H5N1 subtypes (Appendix Table 2) using some amino acid "biomarkers" within each ZF motif.

Of particular importance among the fragments is the motif III (Fig. 2 A & B) which contains residue substitutions that are biologically and functionally relevant in the activities of the protein. This motif is strategically placed at the centre of other motifs and flanked on either side by 2 Histidine residues. It contains non-synonymous amino acid substitutions or mutations between HPAI/LPAI and NPAI suggesting that it can be used to categorize pathogenic and non-pathogenic avian influenza. Interestingly, this is the only motif that contains the classical Cys-His zinc finger (CCHH) and it is the only motif closer to the polybasic HA cleavage site (RRRKK) used in the existing classification of virulence. While the existing classification only differentiate between pathogenic and low/non-pathogenic by molecular method of detecting the presence and absence of HA proteolytic cleavage site sequences, ZF motif III can be useful to further differentiate between low and non-pathogenic (LPAI & NPAI) by analyzing the non-synonymous mutation within the motif of LPAI and NPAI (i.e. red letter residues in Fig. 2B).

The other mutations observed between HPAI and LPAI in amino acid residues of other fragments are synonymous, meaning they can naturally substitute each other without affecting the function of the protein, hence not biologically relevant when compared to that of non-synonymous mutation. The only 2 non-synonymous mutations within ZF motif III occurred at amino acid sequence 291 where Asparagine (N) or Aspartic acid (D) both synonymous in HPAI and LPAI respectively changed to a non-synonymous Isoleucine (I) in NPAI. The other one is the amino acid sequence 315 where Isoleucine in HPAI and LPAI changed to non-synonymous Serine (S) in NPAI. These 2 amino acid substitutions can be utilized to distinguish NPAI from LPAI and/or HPAI. This type of substitution or mutation is an indication that the amino acid interchange in each residue cannot naturally substitute each other without causing functional defect because of lack of functional similarity in their physicochemical properties. In essence, Asparagine (N) does not share any physicochemical properties with Isoleucine (I) and neither does Isoleucine (I) with Serine (S), therefore they cannot replace each other without causing defect in the protein structure and function. Therefore, these 2 residues might be important in determining the virulence of the H5N1 subtypes because a single mutation of I to N or S to I in NPAI may affect the zinc binding activity of the motif and thus cause the non-pathogenic subtype of H5N1 to revert to highly pathogenic or low pathogenic in the presence or absence of cleavage site amino acid sequences respectively.

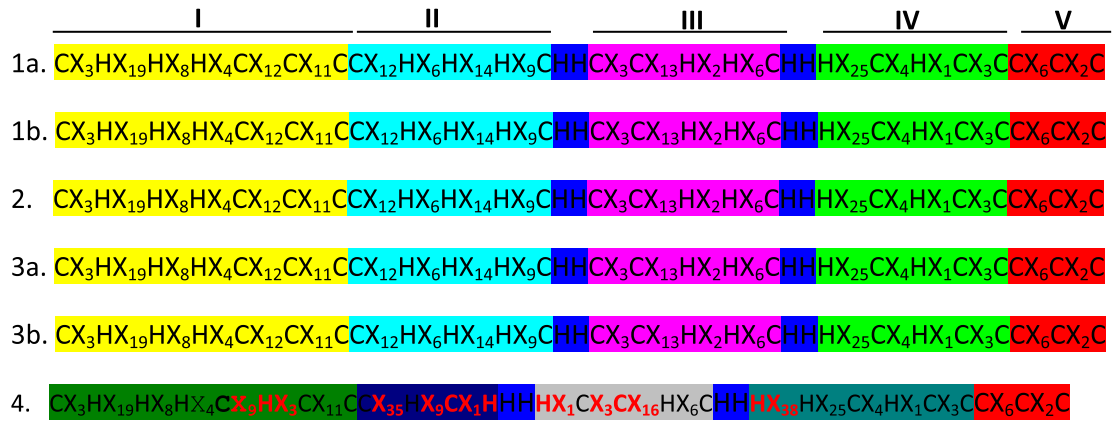

Fig. 1. Full-length sequences of heamaglutinin (HA) protein from different subtypes and lineages of avian influenza viruses highlighting the Cyseine-Histidine (CH)-rich putative 'zinc-binding' motifs shaded in different colours (i.e. motif I-V), separated into 2 parts (I, II) & (IV, V) by 2 molecules of Histidine H (highlighted blue) on either sides of motif III that is centrally located. Note the similarity or full conservation represented by colour coding of each fragment (motif) and the total number of amino acid residues in motifs of same lineage H5 (i.e. subtypes 1, 2 & 3 of H5N1 and H5N2) which differs from another lineage H1 (i.e. subtype 4 of H1N1). C = cysteine, H = histidine, X = any other accompanying amino acids other than C and H. **Keys:** 1a. Highly pathogenic subtype H5N1 (HPAI) of chicken, isolated primarily from chicken in Nigeria (Accession No: CY048083). 1b. Highly pathogenic subtype H5N1 (HPAI) of human, isolated primarily from man in Nigeria (Accession No: EU146920). 2. Non pathogenic subtype H5N1 (NPAI) of duck and chicken, isolated primarily from duck in Japan (accession No: AB530992). 3a. Low pathogenic subtype H5N1 (LPAI) of duck, isolated primarily from black duck in USA (Accession No: ABV25966.1). 3b. Low pathogenic subtype H5N2 (LPAI) of duck and chicken, isolated primarily from duck in Nigeria (Accession No: FR771823). 4. Highly pathogenic subtype H1N1 (HPAI) of man, isolated primarily from man in Nigeria (Accession No: CY073166) NB: 1, 2 & 3a are in the same subtype H5N1 while 3b belong to a different subtype H5N2. 1, 2 & 3 belong to the same lineage H5 while 4 belong to entirely different lineage H1 of subtype H1N1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The other regions where this type of mutation occurs outside the fragments or motifs is at amino acid sequence 336 and 339 after motif III, just before and on the residue of HA cleavage site sequences (PQRE) respectively. Isoleucine (I) in NPAI mutated to a non-synonymous Serine (S) in HPAI which is synonymous to Valine (V) in LPAI. Arginine (R) in NPAI and LPAI mutated to Glycine (G) in HPAI. This particular type cannot be used to distinguish LPAI from NPAI but can differentiate both of them from HPAI and this mutation may be responsible for the insertion of pathogenic cleavage sequences (RRKKR) in HPAI that is currently used in classification of virulence. The only other fragment apart from motif III that contained a non-synonymous amino acid substitution between HPAI and NPAI/LPAI is motif IV at sequence 460 where N mutated to I as in motif III, though this fragment does not have a consensus CCHH zinc finger as that of motif III. Other observed substitution of residues in other fragments such as motif I & II (highlighted brown) are synonymous which naturally may not affect the function of the protein but can equally be used to classify the subtypes as in motif III (Appendix Table 1 and 2).

The analysis above illustrates that there is a distinguishing feature between the highly pathogenic H5N1 subtype of both chicken & humans (i.e. 1a & 1b) from the non-pathogenic H5N1 subtype of chicken (i.e. 2) or low pathogenic H5N2 and H5N1 subtypes of duck/chicken (i.e. 3a & 3b) using "ZF fragments or motifs" (Fig. 1). Interestingly also, there is a residue substitution in the fully conserved cysteine-rich motif V of the different lineages H5 and H1 (Fig. 1), suggesting that the fully conserved residues based on the colour coding between the different lineages can also be distinguished using their amino acid compositions and types (data not shown).

Putting these together, all the 4 residues in motif II can be used to categorize the degree of pathogenicity of the H5N1 subtypes, whereas 3 residues in motif III & IV, and 1 residue in motif I can be used (as indicated by single and double dagger signs in Appendix Table 1). The best of the classification is motif III which is centrally and strategically located containing the consensus CCHH zinc finger and very close to the HA cleavage site.

This new method of virus lineages and subtypes categorization using the 'zinc-finger' motifs has a biological relevant connotation, in that zinc ion is involved in modulating disease pathogenesis. On the other hand, the existing classification of the HPAI subtype of H5N1 uses specific sequences of the 'cleavage site' of haemagglutinin protein in chicken and human (i.e. PQGERRRKKRGLF) [15], or the cleavage sequences of low pathogenic lineage or subtype of H5N2 and H5N1 (LPAI) in duck (i.e. PQRETRGLF) [13] which are not connected to any known biologically relevant activity. Bioinformatics analysis of this cleavage sequence region falls outside the region of any of our potential zinc-binding motifs and as such does not provide any relevant biological correlation of the sequences other than a cleavage site for some enzymes activity in the body.

The relationship of the twenty naturally occurring amino acids with one another in term of their physicochemical properties (e.g. size and hydrophobicity; small, tiny, charged, polar, aromatic or aliphatic) is important in the determination of protein tertiary structure. Some amino acids may share some properties with one another and can therefore substitute one another naturally without resulting in altered function. From our analysis, all the changes observed in each motif have properties to share with each other and can substitute each other without a problem except the 2 residues in motif III (CX_3 & HX_6) identified by double dagger sign for ease of categorization (Appendix Table 1 & 2). This type of mutation may accounts for pathogenicity status of the viral subtypes. Interestingly, the existing method of classification also contains a residue interchange that cannot substitute each other (G and R) between HPAI and LPAI and therefore may equally accounts for introduction of pathogenicity factor in HPAI. Thus, it only takes a mutation of either I to N and S to I (new method) or R to G (existing method) for non-pathogenic subtype to become highly virulence.

A

```
                                                     I
HPAI_HA   MEKIVLLFAIVSLVKSDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDL 60
LPAI_HA   MEKIVLLLTMVSLVKSDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKAHNGKLCSL 60
NPAI_HA   MEKIALLFAIVSLVKSDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCSL 60
          ****.**:::*****************************************:******.*

                                                           II
HPAI_HA   DGVKPLILRDCSVAGWLLGNPMCDEFLNVPEWSYIVEKINPANDLCYPGNFNDYEELKHL 120
LPAI_HA   NGVKPLILRDCSVAGWLLGNPMCDEFLNVPEWSYIVEKDNPVNGLCYPGDFNDYEELKHL 120
NPAI_HA   NGVKPLILRDCSVAGWLLGNPMCDEFLNVPEWSYIVEKDSPINGLCYPGDFNNYEELKHL 120
          :************************************* .* *.*****:**:*******

HPAI_HA   LSRINHFEKIQIIPKSSWSDHEASSGVSSACPYQGRSSFFRNVVWLIKKDNAYPTIKRSY 180
LPAI_HA   LSSTNHFEKIQIIPRSSWSNHDASSGVSSACPYNGRSSFFRNVVWLIKKDNAYPTIKRSY 180
NPAI_HA   LSSTNHFEKIQIIPRSSWSNHDASSGVSSACPYNGRSSFFRNVVWLIKKNNAYPTIKRSY 180
          **  **********:****:*:***********:***************:**********

HPAI_HA   NNTNQEDLLVLWGIHHPNDAAEQTRLYQNPTTYISVGTSTLNQRLVPKIATRSKVNGQSG 240
LPAI_HA   NNTNQEDLLVLWGIHHPNDAAEQTNLYQNPTTYVSVGTSTLNQRSVPEIATRPKVNGQSG 240
NPAI_HA   NNTNQDDLLVLWGIHHPNDAAEQTKLYQNPTTYVSVGTSTLNQRSVPEIATRPKVNGQSG 240
          *****:******************.********:********** **:**** *******
                                                            III
HPAI_HA   RMEFFWTILKPNDAINFESNGNFIAPENAYKIVKKGDSTIMKSELEYGNCNTKCQTPIGA 300
LPAI_HA   RMEFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSGLEYGNCNTKCQTPMGA 300
NPAI_HA   RMEFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSGLEYGNCITKCQTPMGA 300
          *************************** **********:**** ****** ******:**

HPAI_HA   INSSMPFHNIHPLTIGECPKYVKSNRLVLATGLRNSPQGERRRKKRGLFGAIAGFIEGGW 360
LPAI_HA   INSSMPFHNIHPLTIGECPKYVKSDRLVLATGLRNVPQR----ETRGLFGAIAGFIEGGW 356
NPAI_HA   INSSMPFHNIHPLTSGECPKYVKSDRLVLATGLRNIPQR----ETRGLFGAIAGFIEGGW 356
          ************** *********:********** **     :.***************

HPAI_HA   QGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLE 420
LPAI_HA   QGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGITNKVNSIIDKMNTQFEAVGKEFNNLE 416
NPAI_HA   QGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGITNKVNSIIDKMNTQFEAVGKEFNNLE 416
          *********************************:*******************:******
                                                         IV
HPAI_HA   RRIENLNKKMEDGFLDVWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELG 480
LPAI_HA   RRIENLNKKMEDGFLDVWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKEMG 476
NPAI_HA   RRIENLNKKMEDGFLDVWTYNAELLVLMENERTLDFHDSIVKNLYDKVRLQLRDNAKELG 476
          *************************************** ******************:*

HPAI_HA   NGCFEFYHRCDNECMESVRNGTYDYPQYSEEARLKREEISGVKLESIGTYQILSIYSTVA 540
LPAI_HA   NGCFEFYHKCDNECMESVRNGTYDYPQYSEEARLNREEISGVKLESMGTYQILSIYSTVA 536
NPAI_HA   NGCFEFYHKCDDECMESVRNGTYDYPQYSEEARLNREEISGVKLESIGTYQILSIYSTVA 536
          ********:**:**********************:***********:*************
                            V
HPAI_HA   SSLALAIMVAGLSLWMCSNGSLQCRICI       568
LPAI_HA   SSLALAIMIAGLSFWMCSNGSLQCRICI       564
NPAI_HA   SSLALAIMVAGLSFWMCSNGSLQCRICI       564
          ********:****:**************
```

B

```
                          III
HPAI_HA SELEYGNCNTKCQTPIGAINSSMPFHNIHPLTIGECPKYVKSNRLVLATGLRNSPQGERRRKKRTRGLF
LPAI_HA SGLEYGNCDTKCQTPLGAINSSMPFHNVHPLTIGECPKYVKSDKLVLATGLRNVPQRE------TRGLF
NPAI_HA SGLEYGNCITKCQTPMGAINSSMPFHNIHPLTSGECPKYVKSDRLVLATGLRNIPQRE------TRGLF
```

(caption on next page)

Fig. 2. A). Clustal Omega sequence alignment between HPAI, LPAI and NPAI showing the sequences of the 5 putative zinc finger motifs (I-V) shaded in different colours. There are some residue mutations within the motifs which are useful for categorization of pathogenicity of the subtypes. Most of these amino acids mutations or substitutions are synonymous which naturally would not affect the function of the viral HA protein (i.e. the residues highlighted brown). The only exception to this is the mutation with non-synonymous substitution which is likely to affect the function of the viral protein and is found in 2 places within the centrally located motif III (i.e. residues highlighted red) and in 2 places outside this motif just before the HA cleavage site sequences (RRKKR). This type of mutation is also noticed in 1 place in motif IV. Full conservation between amino acid sequences is represented by asterisk sign (*) while partial conservation by single or double dots (. or :) and non-conservation by gap or no sign. **B).** A closer look of the non-synonymous substitutions within and outside motif III which might be responsible for viral virulence as well as distinguishing between the low and non-virulence subtypes that are lacking the polybasic amino acids cleavage site sequences (RRKKR). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Assessment of the 3D protein structures in the PDB database showed that out of the many protein molecules available in the database, only about two structures of H3 among the hemagglutinin protein molecules have been shown to interact with zinc ion ligand at the His residue within Cys residues [16]. The interaction of zinc ion with acidic amino acid side chains and histidines appears to be from the immunoglobulin/receptor and the cysteines are all bound as disulfides. Although there are many 3D molecules of H5 protein in the database [17], their interactions with zinc ion has not been shown in PDB database and this calls for more research on modeling of this protein with zinc ligand to establish the domain of interaction.

In conclusion, since the existing method of classification using the polybasic cleavage sequence domain did not overlap in any of our potential "zinc-finger" sequence domains, it thus makes our method an alternative novel approach of categorization in support of the existing classification. There is possibility of functional importance in our method because of the importance of zinc ion in viral growth and survival, which is mobilized by the specific zinc-finger protein and required for viral attachment, invasion, multiplication and survival in the host. There exist some relatively simple *in vitro* zinc-binding assays to assess the functional importance of the putative zinc finger sequences motif, but to the best of our knowledge there is no such simple assay for the putative HA proteolytic cleavage sequences motif.

Recommendation

Though preliminary, this research hypothesis has opened up opportunity to understand influenza viruses in more details as it concerns zinc ion and zinc binding domain. To further validate the observations in this study, there's a need to test the hypothesis wherein the zinc-binding activity of each of the predicted ZF fragments especially the most important fragment III containing the classical CCHH zinc finger motif is evaluated for zinc functional assay. This assay will be able to tell the distinction between the highly pathogenic avian influenza (HPAI) and non-pathogenic avian influenza (NPAI) of H5N1 subtypes as well as low pathogenic H5N2/H5N1 subtypes in terms of zinc-binding ability and disease causing capability of HA protein.

Methodology for the hypothesis

Biological dataset and information on avian influenza viruses were retrieved from the National Centre for Biotechnology Information (NCBI) database. In higher animals, influenza genus include type A, B, C & D. In birds, type A consists of subtypes H1-18 and N1-11 with many strains within and among the subtypes. The subtypes have a wide host range including avian, tiger, leopard, seal and human among others, with at least 54 countries been infected globally [18–20]. The influenza A viruses have 11 genes on eight separate RNA molecules including the PB2 (polymerase basic 2), PB1 (polymerase basic 1), PB1-F2 (alternate open reading frame near the 5′ end of the PB1 gene), PA (polymerase acidic), HA (hemagglutinin), NP (nucleoprotein), NA (neuraminidase), M1 and M2 (matrix), NS1 (non-structural), NEP/NS2 (nuclear export of vRNPs), with HA and PB1 being the most important for transmissibility and viral polymerase for virulence respectively [21]. The subtyping of avian influenza viruses as well as other influenza viruses are based on two proteins on the surface of the viral envelope viz: the HA (a protein that causes red blood cells to agglutinate) and NA (an enzyme that cleaves the glycosidic bonds of the monosaccharide, neuraminic acid). Based on the encoding of different influenza viruses for the different hemagglutinin and neuraminidase proteins, the virus subtypes are usually named. To date, eighteen H and eleven N subtypes exist in birds and bats [11] indicating that up to 198 different types of influenza viruses may exist.

These molecular structures and the arrangements of the RNA molecules form the basis for a huge and inexhaustible data that can be generated using zinc-finger motifs of avian influenza virus (http://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi#mainform). Because previous studies have confirmed that mutations in the consensus zinc-finger motif (CCHH) caused reduction in the growth and virulence of influenza viruses *in vitro* [5], we obtained full-length sequences of haemagglutinin (HA) protein of avian influenza viruses of H5 subtypes from the repository of avian influenza databases. The viruses included in the analysis were the highly pathogenic avian influenza H5N1 (ten sequence data), non-pathogenic avian influenza H5N1 (two sequence data), low pathogenic avian influenza H5N1 (two sequence data) as well as low pathogenic H5N2 (ten sequence data) from different avian species including chicken, waterfowls and human from Africa and other continents. We manually mapped out the conserved putative Cys-His (CH)-rich region proposed to represent a potential zinc-binding domains in different subtypes of HA sequences, then compared the number and types of amino acid residues in each of them to identify any unique distinguishing feature(s) between the subtypes which can be used as a potential biomarker for viral categorization and/or pathogenicity. We showed the similarity or conservativeness in total numbers and compositions of residues in the zinc finger motifs of H5 viral lineage (e.g. H5N1 & H5N2) and how this differed from other viral lineage such as the H1 (e.g. H1N1) viruses.

Acknowledgement

The authors wish to thank Prof Wolfgang Maret of King's College London for his insight on 3D structures.

Appendix

Table 1
Result of analysis of the changes in the composition and type of amino acids within each ZF motif between the HPAI and NPAI of H5N1 subtypes (i.e. 1 & 2) using 1a as a reference for comparison.

Motif	Residue	Changes in amino acids & position of change in the residue	Subtype with changes differing from 1a	Subtype with changes similar to 1a
I	CX_3	None	None	All (HPAI & NPAI)
	HX_{19}	None	None	All (H & NPAI)
	HX_8	None	None	All (H & NPAI)
	HX_4	None	None	All (H & NPAI)
	CX_{12} †	D^1S, D^3N	NPAI (2)	HPAI (1a & 1b)
	CX_{11}	None	None	All (H & NPAI)
II	CX_{12} †	N^4D, D^7N	NPAI (2)	HPAI (1a & 1b)
	HX_6 †	R^4S I^5T	NPAI (2)	HPAI (1a & 1b)
	HX_{14} †	K^9R, $D^{14}N$	NPAI (2)	HPAI (1a & 1b)
	HX_9 †	E^1D	NPAI (2)	HPAI (1a & 1b)
III	CX_3 ‡	N^1I	NPAI (2)	HPAI (1a & 1b)
	CX_{13} †	I^4M	NPAI (2)	HPAI (1a & 1b)
	HX_2	None	None	All (H & NPAI)
	HX_6 ‡	I^4S	NPAI (2)	HPAI (1a & 1b)
IV	HX_{25} ‡	N^3I	NPAI (2)	HPAI (1a & 1b)
	CX_4	None	None	All (H & NPAI)
	HX_1 †	R^1K	NPAI (2)	HPAI (1a & 1b)
	CX_3C †	N^2D	NPAI (2)	HPAI (1a & 1b)
V	CX_6	None	None	All (H & NPAI)
	CX_2	None	None	All (H & NPAI)

Analysis of changes in sequence of amino acid residues that can serve as biomarkers for viral categorization into highly pathogenic H5N1 and non-pathogenic H5N1 subtypes which are represented by the single dagger sign (†) with the subtypes indicated in red. The residue changes with double daggers sign (‡) are in addition useful for categorization into highly pathogenic and non-pathogenic H5N1 subtypes.

Table 2
Result of analysis of the changes in the composition and type of amino acids residues between the HPAI of H5N1 and LPAI of H5N2/H5N1 subtypes (i.e. 1 & 3) using 1a as a reference for comparison.

Motif	Residue	Changes in amino acids & position of change in the residue	Subtype with changes differing from 1a	Subtype with changes similar to 1a
I	CX_3 HX_{19} HX_8†* HX_4 CX_{12}† CX_{11}	None None T^8A None D^1S, D^3N None	None None LPAI & HPAI (3 & 1b) None LPAI (3) None	All (HPAI & LPAI) All (HPAI & LPAI) HPAI (1a) All (HPAI & LPAI) HPAI (1a & 1b) All (HPAI & LPAI)
II	CX_{12} † HX_6 † HX_{14} † HX_9 †	N^4D R^4S K^9R, $D^{14}N$ E^1D	LPAI (3) LPAI (3) LPAI (3) LPAI (3)	HPAI (1a & 1b) HPAI (1a & 1b) HPAI (1a & 1b) HPAI (1a & 1b)
III	CX_3 CX_{13}† HX_2 HX_6	None I^4M None None	None LPAI (3) None None	All (HPAI &LPAI) HPAI (1a & 1b) All (HPAI & LPAI) All (HPAI &LPAI)
IV	HX_{25}†* CX_4 HX_1 † CX_3C	$L^{22}M$ None R^1K None	LPAI & HPAI (3 & 1b) None LPAI (3) None	HPAI (1a) All (HPAI & LPAI) HPAI (1a & 1b) All (HPAI &LPAI)
V	CX_6 CX_2	None None	None None	All (HPAI & LPAI) All (HPAI & LPAI)

Analysis of changes in sequence of amino acid residues that can serve as biomarkers for viral categorization into highly pathogenic H5N1 and low-pathogenic H5N2 subtypes which are represented by the single dagger sign (†) with the subtypes indicated in red. It is interesting to note the same amino acids residue changes that are useful for categorization into highly pathogenic and non-pathogenic H5N1 subtypes (in Table 1) are also useful for categorization into highly pathogenic H5N1 and low pathogenic H5N2/H5N1 subtypes (in Table 2) (†), exception are the residue changes with double daggers sign (‡) only present in Table 1. Therefore, only the analysis of residues changes with double daggers sign (‡) are important to differentiate between the subtypes of non-pathogenic H5N1 and low pathogenic H5N2/H5N1 that are both lacking HA polybasic cleavage sites. The residue changes represented by single dagger signs with asterisk (†*) which does not distinguish between HPAI and NPAI may be useful to differentiate HPAI of birds from humans.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mehy.2020.109925.

References

[1] Gammoh NZ, Rink L. Zinc in Infection and Inflammation. Nutrients 2017;9(6):624.
[2] Liu T, Ye Z. Restriction of viral replication by mutation of the influenza virus matrix protein. J Virol 2002;76(24):13055–61.
[3] Ebert JC, Altman RB. Robust recognition of zinc binding sites in proteins. Protein Sci 2008;17(1):54–65.
[4] Tummino PJ, Scholten JD, Harvey PJ, Holler TP, Maloney L, Gogliotti R, et al. The *in vitro* ejection of zinc from human immunodeficiency virus (HIV) type 1 nucleocapsid protein by disulfide benzamides with cellular anti-HIV activity. Proc Natl Acad Sci 1996;93:969–73.
[5] Hui EK, Smee DF, Wong M, Nayak DP. Mutation in influenza virus M1 CCHH, the putative zinc-finger motif, cause attenuation in mice and protect mice against lethal influenza virus infection. J Virol 2006;80(12):5697–707.
[6] Huang C, Kiyotani K, Fujii Y, Fukuhara N, Kato A, Nagai Y, et al. Involvement of the zinc-binding capacity of Sendai virus V protein in viral pathogenesis. J Virol 2000;74:7834–41.
[7] Modrof J, Becker S, Muhlberger E. Ebola virus transcription activator VP30 is a zinc-binding protein. J Virol 2003;77(5):3334–8.
[8] Steward M, Samson A, Errington W, Emmerson PT. The Newcastle disease virus V protein binds zinc. Arch Virol 1995;140:1321–8.
[9] Seybert A, Posthuma CC, van Dinten LC, Snijder EJ, Gorbalenya AE, Ziebuhr J. A complex zinc finger controls the enzymatic activities of nidovirus helicase. J Virol 2005;79(2):696–704.
[10] Lycett SJ, Ward MJ, Lewis FI, Poon AFY, Kosakovsky Pond SL, Leigh Brown AJ. Detection of mammalian virulence determinants in highly pathogenic avian influenza H5N1 viruses: multivariate analysis of published data. J Virol 2009;83(19):9901–10.
[11] Tong S, Zhu X, Li Y, Shi M, Zhang J, Bourgeois M, et al. New world bats harbor diverse influenza A viruses. PLoS Pathog 2013;9(10):e1003657.
[12] Alexander DJ. A review of avian influenza in different bird species. Vet Microbiol 2000;74:3–13.
[13] Coker T, Meseko C, Odaibo G, Olaleye D. Circulation of the low pathogenic avian influenza subtype H5N2 virus in ducks at a live bird market in Ibadan. Nigeria Inf

Dis Pov 2014;3:38.
[14] Luczo JM, Tachedjian M, Harper JA, Payne JS, Butler JM, Sapats SI, et al. Evolution of high pathogenicity of H5 avian influenza virus: haemagglutinin cleavage site selection of reverse-genetics mutants during passage in chickens. Sci Rep 2018;8(1):11518.
[15] Toshihiro I, Hideo G, Eiji Y, Hiroko T, Mutsuko T, Masaru K, et al. Generation of a highly pathogenic avian influenza A virus from an avirulent field isolate by passaging in chickens. J Virol 2001;75(9):4439–43.
[16] Fleury D, Wharton SA, Skehel JJ, Knossow M, Bizebard T. Antigen distortion allows influenza virus to escape neutralization. Nat Struct Biol 1998;5:119–23.
[17] Xiong X, Coombs PR, Martin S, Liu J, Xiao H, Mccauley JW, et al. Receptor binding by a ferret-transmissible H5 avian influenza virus. Nature 2013;497:392.
[18] Subbarao K, Klimov A, Katz J, Regnery H, Lim W, Hall Henrietta, et al. Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. Science 1998;279(5349):393–6.
[19] Keawcharoen J, Oraveerakul K, Kuiken T, Fouchier RAM, Amonsin A, Payungporn S, et al. Avian influenza H5N1 in tigers and leopards. Emerg Infect Dis 2004;10:2189–91.
[20] OIE. Update on avian influenza in animals (types H5 and H7). Available on: https://www.oie.int/en/animal-health-in-the-world/update-on-avian-influenza/. Accessed 29 March 2020.
[21] Bouvier NM, Palese P. The biology of influenza viruses. Vaccine 2008;26(4):49–53. https://doi.org/10.1016/j.vaccine.2008.07.039.