Journal of Molecular Graphics and Modelling 79 (2018) 185–191

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

Structural comparison among the 2013–2017 avian influenza A H5N6 hemagglutinin proteins: A computational study with epidemiological implications

Marni E. Cueno[a,b,*], Izuho Suzuki[b], Shiori Shimotomai[b], Takuma Yokoyama[b], Kai Nagahisa[b], Kenichi Imai[a]

[a] Department of Microbiology, Nihon University School of Dentistry, Tokyo 101-8310, Japan
[b] Immersion Physics Class, Department of Science, Tokyo Gakugei University International Secondary School, Tokyo 178-0063, Japan

ARTICLE INFO

Article history:
Received 27 July 2017
Received in revised form
25 November 2017
Accepted 27 November 2017
Available online 28 November 2017

Keywords:
Avian influenza
Hemagglutinin
Homology modeling
Influenza a H5N6 subtype
Structural evolution

ABSTRACT

Avian influenza viruses easily spread allowing viral re-assortment to simply occur which in-turn increases the potential for a pandemic. A novel 2013 H5N6 influenza strain was detected among the avian population and was reported to continuously evolve, however, this was never structurally demonstrated. Here, we elucidated the putative structural evolution of the novel H5N6 influenza strain. Throughout this study, we analyzed 2013–2017 H5N6 HA protein models. Model quality was first verified before further analyses and structural comparison was made using superimposition. We found that Leu was inserted at position 129_1 among the 2013–2015 models while Leu was not inserted among the 2016–2017 models. Moreover, presence of Leu at position 129_1 shifts residue $E126_1$ by 159.6° affecting nearby residues which may explain the difference between the 2013–2015 and 2016–2017 HA structural groups. Similarly, we believe that our results would support the hypothesis that the current H5N6 strain is still continuously evolving.

1. Introduction

Influenza A virus is a type of RNA virus that utilizes sialic acid binding to the host cell surface in order to instigate an infection [1]. Hemagglutinin (HA) is a homotrimeric glycoprotein that constitutes most of the virus surface and is the main viral protein structure involved in sialic acid binding to the host cell [2]. Each HA can be divided into two polypeptides generated from a single nascent peptide chain through protease cleavage, namely: HA1 and HA2 [2,3]. HA1 is the membrane-distal domain that is further subdivided into the receptor-binding and vestigial esterase subdomains. On the other hand, the F fusion subdomain (HA2) and both the N- and C-terminal segments of an F' fusion subdomain (HA1) comprise the stem region [4]. HA binding to sialic acid found in the host surface determines viral infectivity which in-turn would serve as a major determinant on what host can be infected [5]. This highlights the functional significance of HA for the influenza virus since it is involved in viral transmissibility and evolution [1,5,6]. Additionally, HA binding properties are crucial to determine influenza evolution and is influenced by several factors such as inter-residue atomic interactions [7].

Avian influenza viruses (AIV) are easily spread due to undomesticated birds and, upon infecting an avian host, the virus is localized and replicates in the host intestinal tract which then is excreted in high amounts in the faeces and, subsequently, may orally be ingested by a potential avian host [8]. This would suggest that faecal-oral transmission route allows for viral re-assortment to occur easily and continuously within the avian population and, more importantly, increases the potential to infect humans [9–11]. In 2013, a novel H5N6 influenza strain was detected among the avian population and has been reported to continuously evolve [8,11,12]. However, this was never structurally elucidated. A better understanding on how the novel H5N6 influenza strain evolved during 2013–2017 may help shed light on the extent of viral evolution, thereby, could help predict the occurrence of future epidemic or pandemic strains which, in-turn, could lead to novel antiviral strategies.

* Corresponding author at: Department of Microbiology, Nihon University School of Dentistry, 1-8-13 Kanda-Surugadai, Chiyoda-ku, Tokyo 101-8310, Japan.
E-mail address: marni.cueno@nihon-u.ac.jp (M.E. Cueno).

https://doi.org/10.1016/j.jmgm.2017.11.013
1093-3263/

2. Methodology

2.1. Protein modeling

H5N6 HA amino acid sequences from 2013 to 2017 influenza strains were collected from the National Center for Biological Information (NCBI) website. Representative strains used were the following: 2013 strain with Genebank accession number AJS16152.1; 2014 strain with Genebank accession number AJS16157.1; 2015 strain with Genebank accession number APG39822.1; 2016 strain with Genebank accession number BAW32343.1; and 2017 strain with Genebank accession number AQU11868.1. Representative crystal structure used for superimposition [13] was the 2014 strain (PDB ID: 5HU8; PMID: 27053557). We followed the numbering scheme as previously published [14] and we designated subscript 1 and 2 to refer to HA_1 and HA_2, respectively.

HA protein models were generated using the Phyre2 web server [15]. Briefly, the Phyre web server utilizes known protein structures from both SCOP and PDB databases and, subsequently, each predicted 3D protein structure is scanned against a non-redundant sequence database while each user-submitted sequence is scanned against the non-redundant sequence database. Finally, a possible protein model profile is constructed with the query secondary structure being predicted by three independent secondary structure prediction programs (Psi-Pred [16], SSPro [17], JNet [18]) and the confidence values of each program are averaged and consensus calculated. Moreover, both protein model profile and putative secondary structures are scanned against a pre-existing fold library using a profile–profile algorithm [19] which in-turn returns an alignment ranking score that is fitted to an extreme value distribution in order to generate an *E*-value, whereby, the top ten highest scoring alignments are then used to construct full 3D models. All protein models were visualized using the Jmol applet [20].

2.2. Model quality verification

Protein model quality was determined through: molecular dynamics simulation, model quality estimation, and protein model:crystal structure superimposition. Coarse grain-molecular dynamics (CG-MD) simulation was performed to determine the radius of gyration (R_{gyr}) of the generated HA models using the MDWeb server [21] which is based on the original MDMoby software platform that makes use of the Amber tools and VMD packages. Additionally, CG-MD simulation conditions were set at 1000 ps simulation time with Δt at 0.01 ps and output frequency collected at 10 ps. Briefly, protein models that have minimal R_{gyr} observed are considered stable and are structurally reliable. Moreover, Qualitative Model Energy Analyses (QMEAN) scores were determined to estimate the quality of each protein model generated [22,23]. Briefly, QMEAN score is based on the linear combination of six structural descriptors and reflects the predicted global model reliability ranging from 0 to 1, wherein, scores close to 1 are considered reliable. For this study, as an additional model quality verification parameter, QMEAN scores between a crystal structure and protein model were compared, wherein, protein model QMEAN scores that are close to the crystal structure QMEAN scores would mean to indicate that the protein model can be considered reliable. Furthermore, generated HA protein models were superimposed with a known H5N6 HA crystal structure using SuperPose [13] in order to establish the Root Mean Square Deviation (RMSD) values of the superimposed Cα backbone. For purposes of model quality verification, RMSD values <1.5 Å were considered reliable.

2.3. Structural comparison

For this study, 2 sets of protein model comparison were performed: (1) comparison between the different HA protein models generated from different H5N6 strains; and (2) comparison between the original and mutated versions of the different HA protein models generated from both the same and different H5N6 strains. All protein model comparison done throughout the study was performed using SuperPose and the computed RMSD values generated during superimposition were determined. For protein model comparison, RMSD values close to 0 were considered similar.

Mutated versions of the HA protein models were generated by either inserting or deleting a candidate amino acid residue that was judged to have the potential to significantly affect certain structural properties which in-turn could result in protein structure alteration. Subsequently, structural properties that were altered after either inserting or deleting candidate amino acid residue/s were likewise elucidated.

3. Results and discussion

3.1. Generated HA protein models are accurate and reliable

Protein models produced either experimentally or theoretically (computer-based) should be assessed for accuracy and reliability before performing further analyses [24]. To confirm the accuracy and reliability of each HA protein model generated, three different strategies were done, namely: CG-MD simulation, QMEAN scoring, and superimposition. For this study, HA monomers were used for both modeling and validation which in-turn limited our analyses to HA structural differences within the HA protein model. It is worth mentioning that using HA trimers for structural analyses would have factored in protein–protein interactions among the HA monomers which consequently could influence the HA structure and, more importantly, would affect our structural analyses. As seen in Fig. 1A, all generated protein models (*left panel*) were found to have minimal R_{gyr} (*right panel*). Similarly, both crystal structure and protein model QMEAN scores were close to 1 and are relatively close to one another (Fig. 1B, *left and middle panels*). Moreover, superimposition of a crystal structure and representative protein model showed an RMSD value <1.5 Å (Fig. 1B, *right panel*). Overall, these results would mean that the generated HA protein models are potentially similar with that of the crystal structure and, more importantly, can be used for further downstream structural analyses.

3.2. HA protein models differ between the 2013–2015 and 2016–2017 H5N6 strains

The AIV pool is comprised of highly pathogenic avian influenza virus (HPAIV) and low pathogenic avian influenza virus (LPAIV) populations. HPAIV has been attributed to infections causing highly contagious systemic diseases with high morbidity and mortality to both avian and human populations, whereas, LPAIV trigger milder respiratory diseases mostly in the avian population [12]. Both HPAIV and LPAIV populations co-circulate in nature which in-turn could allow accumulation of site mutation, genetic reassortment, and genetic recombination among the two populations giving rise to novel AIV strains quickly and frequently [25]. Moreover, in the case of the novel H5N6 influenza strain, it has been proposed that this strain is still continuously evolving [11,12]. This would suggest that the HA protein may potentially differ among the 2013–2017 H5N6 strains. To determine any possible HA structural differences among the 2013–2017 H5N6 strains, superimposition between the HA protein models was performed and the RMSD val-