Archives of Biochemistry and Biophysics 625-626 (2017) 13–16

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

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

A singular mutation in the hemagglutinin of the 1918 pandemic virus

Yves-Henri Sanejouand

UFIP, UMR 6286 of CNRS, Faculté des Sciences et des Techniques, Nantes, France

ARTICLE INFO

Article history:
Received 25 April 2017
Received in revised form
19 May 2017
Accepted 24 May 2017
Available online 31 May 2017

Keywords:
Influenza
Pandemic
Hemagglutinin
Mutations

ABSTRACT

The influenza pandemic of 1918–1919 killed at least 50 million people. The reasons why this pandemic was so deadly remain largely unknown [9]. However, It has been shown that the 1918 viral hemagglutinin allows to reproduce the hallmarks of the illness observed during the original pandemic [11]. Thanks to the wealth of hemagglutinin sequences accumulated over the last decades, amino-acid substitutions that are found in the 1918–1919 sequences but rare otherwise can be identified with high confidence. Noteworthy, Gly 188, which is located within a key motif of the receptor binding site, has never been observed again in sequences of human viruses of subtype H1. Monitoring this singular mutation in viral sequences may help prevent another dramatic pandemic.

1. Introduction

It was recently shown that introducing the 1918 hemagglutinin (HA) confers enhanced pathogenicity in mice to recent human viruses that are otherwise non-pathogenic in this host. Moreover, like the 1918 one, these recombinant viruses infect the entire lung and induce high levels of macrophage-derived chemokines and cytokines, which results in infiltration of inflammatory cells and severe haemorrhage [11]. In macaques, the whole 1918 virus causes a highly pathogenic respiratory infection that culminates in acute respiratory distress and a fatal outcome [10]. Although the 1918 polymerase genes were found essential for optimal virulence, replacing the 1918 HA by a contemporary human one proved enough for abolishing the lethal outcome of the 1918 virus infection in mices [23], further underlining the key role of the 1918 HA in the deadly process.

The first 1918 HA sequences were obtained in 1999 from formalin-fixed, paraffin-embedded lung tissue samples prepared during the autopsy of victims of the influenza pandemic, as well as from a frozen sample obtained by *in situ* biopsy of the lung of a victim buried in permafrost since 1918 [17]. Since then, the number of HA sequences determined each year has grown dramatically, jumping from ≈ 100 per year in the eighties to ≈ 3500 per year nowadays.[1] The goal of the present work is to take advantage of this wealth of data for identifying features that are unique to the 1918 sequence, the underlying hypothesis being that they may prove responsible for the unique behaviour of viruses displaying the 1918 HA.

2. Methods

52005 HA protein sequences, of all known subtypes (H1 to H18), were retrieved[2] from the NCBI influenza virus resource [2], sequences coming from laboratory viral strains being disregarded. Note that the redundancy of this sequence dataset is low (≈ 20%). Since the multiple sequence alignment of large numbers of sequences is challenging [4,19,22], and because HA sequences have fair levels of sequence identities, being at least 35% identical to each other, pairwise alignments were performed with BLAST [1], taking as a query the sequence of strain A/Thailand/CU-MV10/2010 (accession number in Genbank: HM752477), one of the longest H1 sequences known. MVIEW [3], version 1.60.1, was then used for converting the BLAST output into an actual multiple sequence alignment (MSA). In this MSA, 3365 different amino-acid residues are observed at least 20 times in 541 different positions (sites), that is, 6.2 per site, on average.[3] Since 1999, eleven 1918–1919 HA sequences have been determined, at least partially. They differ at most by a couple of mutations [17,18]. The sequence of strain A/South Carolina/1/1918 was chosen as a representative. For the HA of this strain, ten X-ray structures are known. In all but one, the residue numbering comes from the H3 subtype. It is also used herein, as provided in PDB structure 4EEF [24].

E-mail address: yves-henri.sanejouand@univ-nantes.fr.

[1] According to the NCBI influenza virus resource.

[2] On September 6 th, 2016.

[3] There are 575 sites in this alignment.

http://dx.doi.org/10.1016/j.abb.2017.05.013
0003-9861/

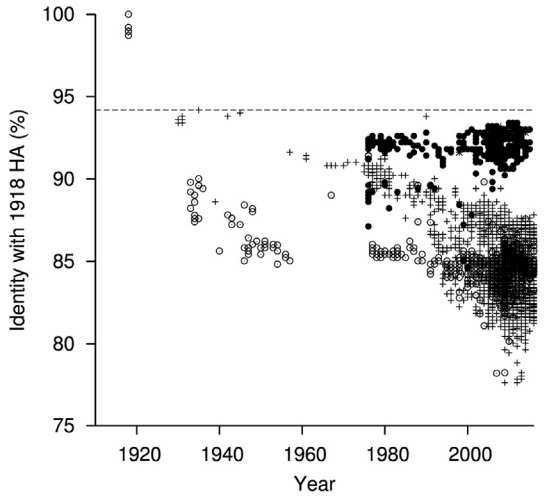

Fig. 1. *Percentage of identity of influenza A hemagglutinin sequences with the 1918 sequence, as a function of time.* The year comes from the collection date provided by the NCBI influenza virus resource. The dotted line indicates the highest level of sequence identity observed since 1918 (94.2%). Viruses come from human (open circles), swine (plus symbols), avian (filled circles) or other hosts (stars). Results are only shown for the H1 subtype, since other subtypes have sequence identities with the 1918 sequence that are below 75%.

Table 1
Number of post-1919 hemagglutinin sequences with same residue as the 1918 sequence. Top: residues found in less than 50 human H1 sequences. Bottom: key residues involved in receptor binding. Bold: residue index of a highly conserved residue, that is, a residue found in more than 95% of the H1 sequences.

1918 HA residue[a]	Human H1	Avian H1	Swine H1	All H1	All HA
Gly 188[c]	0	0	47	47	82
Arg **116**[d]	4	438	162	621	2071
Ala 147[d]	13	0	91	106	8902
Ser 159[c]	17	21	47	85	4066
His 285[c]	17	433	802	1269	1297
Ala 219[c]	22	458	1642	2141	5964
Asn 72[d]	22	468	289	796	7489
Tyr 141[c]	27	458	1341	1850	8222
Lys **120**[c]	29	446	164	656	10549
Arg **153**[d]	31	39	418	488	22855
Ser 264[c]	45	463	3562	4090	9738
Leu 70[c]	48	439	2857	3367	25266
Asp 190[c]	10055	24	5340	15436	28495
Ser 193[c]	7238	312	2758	10334	15936
Lys **222**[c]	10871	432	6101	17437	23284
Asp 225[c]	9940	12	4183	14150	18093
Total	11062	508	6183	17794	51992[b]

[a] H3 subtype residue numbering.
[b] The total number of post-1919 HA sequences.
[c] HA1 subunit.
[d] HA2 subunit.

3. Results

As shown in Fig. 1, known HA sequences of post-1919 viruses are *all*[4] less than 95% identical to the 1918 one, that is, complete ones differ from the 1918 sequence by more than 25 amino-acid substitutions. Strinkingly, after 1919, HA sequences of human viruses are less than 92% identical to the 1918 HA, a ≈ 90% level of sequence identity being observed in the thirties as well as more recently, though in a few instances only.[5] On the other hand, HA sequences of avian viruses more than 93% identical to the 1918 HA have been observed each year since 2005, suggesting that there is a selection pressure favoring 1918 HA-like sequences in avian species, in line with the hypothesis of an avian origin for the 1918–1919 pandemic [17]. Moreover, among the 41 post-1950 sequences that are more than 93% identical to the 1918 HA, 35 (85%) come from duck species,[6] further pinpointing aquatic birds as a possible reservoir [7,12].

However, the 1918 HA sequence is a singular one. For instance, nearly 20 amino-acid residues in this sequence are found in less than 1% of post-1919 human H1 sequences. The twelve less frequent ones are shown on top of Table 1. Most of them are either frequent in avian H1 sequences, or often found[7] in sequences of other HA subtypes (last column).

As a striking exception, after 1919, Gly 188 has *not* been observed again in human H1 sequences. It has also *not* been observed in avian ones. As a matter of fact, it has only been observed in 47 H1 sequences, all of them from swine, a single time in 2003, once each year between 2009 and 2012, and several times each year since then.[8] In human HA of other subtypes, it has been observed 11 times, in sequences from H3N2 or H5N1 viruses. Overall, since 1919, Gly 188 has *only* been observed 82 times, the first time in 2000, in the sequence of an avian H9 HA. Compared to the other residues of the 1918 sequence, Gly 188 is atypical (see Fig. 2) since most residues of the 1918 sequence[9] have been observed in at least 10000 post-1919 HA sequences.

Residue 188 is located at the N-terminus of the 190-helix, which is involved in the HA receptor binding site. However, it does not interact directly with the receptor [5,20]. Three residues are usually observed at this position, namely, serine (41%), isoleucine (33%) and threonine (21%). Interestingly, proline which, like glycine, can have a direct impact on the secondary structure [8], the folding [13] or the stability [15] of a protein, is also rarely observed there[10]. This suggests that the fitness of the virus is impaired when the residue at position 188 is either too flexible (glycine) or when its conformation is atypical (proline). In other words, taken alone, Gly 188 is expected to be deleterious.

However, since the 1918 virus proved efficient, at least one compensatory mutation has to be present in its HA sequence, a phenomenon for instance observed in the HA of the A(H1N1) pdm09 strain [14]. Possible candidates are mutations of residues that are otherwise highly conserved in H1 sequences, like R120K, in the HA1 subunit, K116R or K153R, in the HA2 one (Table 1). Other candidates are 1918 residues that have also been rarely observed since then (Table 1). Though it is not clear how a mutation in the signal peptide could balance the effect of a mutation nearby the binding site, it has to be mentioned that in the fourth position of the signal peptide of the 1918 HA there is an arginine that has been observed there only 59 times, that is, less frequently than Gly 188 itself. Another possibility is the I70L mutation, in the HA1 subunit, which has only been observed 48 times in post-1919 human sequences (Table 1) but has been found in 90% of the 82 HA sequences with Gly 188. Although an isoleucine to leucine substitution is expected to have a less stringent effect than the introduction of a glycine in a protein, it has already been shown that an isoleucine to

[4] For sequence identity comparisons, 14738 short sequences were disregarded, namely, those with less than 400 amino-acid residues.
[5] Four cases in 1976, for instance.
[6] Among the six remaining ones, two come from other avian species, goose and sparrow, and three from swine; the last one was found in the environment.
[7] At least 1297 times.
[8] 11 times in 2014, 14 times in 2015.
[9] 92% of them.
[10] 10 times in all HA sequences.