Virus Genes 24:1, 57–63, 2002

Sequence Similarities and Evolutionary Relationships of Influenza Virus A Hemagglutinins

ANDREA KOVÁČOVÁ,[1,2] GABRIEL RUTTKAY-NEDECKÝ,[1]
IVAN KAROL HAVERLÍK[1] & ŠTEFAN JANEČEK[2,*]

[1]*Department of Biophysics and Chemical Physics, Faculty of Mathematics, Physics and Informatics, Comenius University, Mlynská dolina, SK-84248, Bratislava, Slovakia;*
[2]*Institute of Molecular Biology, Slovak Academy of Sciences, Dúbravská cesta 21, SK-84251, Bratislava, Slovakia*

Received July 19, 2001; Accepted September 12, 2001

Abstract. This study brings the analysis of amino acid sequences of hemagglutinin (HA) from the influenza virus A that can infect a wide variety of birds and mammals. 191 sequences belonging to all known 15 HA subtypes were compared. The emphasis was given on functional sites (receptor-binding cavity with its right and left edges) and degree of their conservation in each subtype. Three evolutionary trees of 15 avian HA representatives were constructed: one tree based on the alignment of the entire HA sequences and two trees based on the alignment of HA1 and HA2 chains, respectively. The results have shown that, despite low degree of sequence similarity among the 191 sequences of HA1 subunit, the active site is well conserved, and that there are only marginal differences in the clustering of the individual HA subtypes between the two subunit trees. In this respect, the subtype H9 seems to be the most fluctuating example. The proposals of the probable avian HAs that could be the closest relatives to human (mammalian) HAs were also provided for several HA subtypes.

Key words: evolution, hemagglutinin, influenza virus A, sequence similarities

Introduction

Influenza viruses can infect a wide variety of birds and mammals. Natural reservoir of the virus is thought to be wild waterfowl. The other animal species (e.g. chickens, turkeys, pigs, horses) including humans are infected with influenza viruses as aberrant hosts [1]. Of the three influenza viruses A, B and C, only certain subtypes of influenza A and the type B cause diseases in humans [2]. During the influenza pandemic in 1918–19, when the human population acquired the virus from birds through the swine, the virus type A (A/South Carolina/1/18; H1N1) caused the death of over 20 million people [3,4].

Influenza virus type A belongs to the family *Orthomyxoviridae*. It is an enveloped negative-strand RNA virus with single stranded segmented RNA genome consisting of eight segments [5]. The viral envelope contains anchored two surface glycoproteins: hemagglutinin (HA) and neuraminidase [6].

HA is a trimer-forming glycoprotein [7] performing two crucial functions: (i) binding of the virus to the surface receptors of the host cell through the sialic acid; and (ii) releasing of the viral genome into the cytoplasm during the process of membrane fusion [8,9]. HA is synthesised in endoplasmic reticulum as a single protein. To be activated, i.e., render the virus infective and fusogenic, its cleavage is necessary by the specific host cell proteases into two chains (subunits) HA1 a HA2 [10].

The influenza virus A undergoes a continuous selection pressure with emerging new variants or reassortants with changed antigenicity of HA that are not recognised by the immune system of the host. These processes may result in pandemics which can

*Author for all correspondence
E-mail: umikstef@savba.sk

be a consequence of a novel virus subtype of influenza A created by reassortment of the segmented genome, and epidemics which appears after the evolution of the surface antigens of influenza A virus [2].

There are 15 different subtypes of HA known. Amino acid sequence comparison revealed some residues involved in, e.g., receptor specificity and sensitivity to neutralisation by a HA inhibitor [11,12]. The evolutionary relationships among the individual HA subtypes were analysed by comparison of either nucleotide or amino acid sequences [13–16]. Until now the phylogenetic tree for all 15 HA subtypes of the influenza A virus based on the comparison of amino acid sequences has not been published. In this study, we focus on all available HA amino acid sequences in an effort to: (i) compare their sequences with regard to conservative functional sites; (ii) find out the degree of sequence conservation in the frame of each of the 15 HA subtypes; (iii) compare the phylogenetic trees based on the alignments of the precursor chain HA and the individual processed chains HA1 and HA2; and (iv) contribute to the whole picture of the influenza virus A evolution.

Materials and Methods

The amino acid sequences of the influenza virus type A HAs were retrieved from GenBank [17]. Since birds have been the natural reservoir of the virus we focus on the HAs isolated from avian viruses (Table 1). Based on the HA sequences, there are 15 different subtypes (H1–H15) of the influenza virus A. The other criterion in selection of the 15 HA subtype representatives was the date of revealing of the virus isolate.

Multiple amino acid sequence alignment was performed on the prepared set of 15 different subtypes of avian HAs using the program CLUSTAL W [18] and then manually tuned where applicable. If there were at least two different sequences for an individual subtype, their subtype alignments were also made. The method used for building the evolutionary tree was the neighbour-joining method [19]. The Phylip format tree output was applied using the bootstrapping procedure [20]; the number of bootstrap trials used was 1000. The tree was drawn with program TreeView [21].

Results and Discussion

In this study, 191 amino acid sequences of influenza virus A HAs belonging to 15 different subtypes were compared (Table 2). There was only one sequence available for subtypes H8, H11, H12 and H14. As can be seen from Table 2, the HA sequences are well conserved since the consensus length for subtype alignments was in each case close to 550 residues. This fact is reflected also by the values of sequence identity and similarity which was not lower than 50% and 66%, respectively, even if more than 50 representatives of a given subtype (H1) were

Table 1. Representative sample of HAs

Subtype	Source	GenBank	GenPept
H1	Duck/Alberta/35/76/H1N1	AF091309	AAD25304.1
H2	Mallard/MT/Y61/H2N2	L11136	AAA43345.1
H3	Duck/Memphis/928/74/H3N8	M73722	—
H4	Duck/Czechoslovakia/56/H4N6	D90302	BAA14332.1
H5	Chicken/Scotland/59/H5N1	X07826	CAA30680.1
H6	Shearwater/Australia/1/72/H6N5	D90303	BAA14333.1
H7	FPV/Rostock/34/H7N1	M24457	AAA43150.1
H8	Turkey/Ontario/6118/68/H8N4	D90304	BAA14334.1
H9	Turkey/Wisconsin/66/H9N2	D90305	BAA14335.1
H10	Chick/Germany/N/49/H10N7	M21647	AAA79775.1
H11	Duck/England/56/H11N6	D90306	BAA14336.1
H12	Duck/Alberta/60/76/H12N5	D90307	BAA14337.1
H13	Gull/Maryland/704/77/H13N6	D90308	BAA14338.1
H14	Mallard/Gurjev/263/82/H14	M35997	—
H15	Shearwater/Australia/2576/79/H15N9	L43917	AAA96134.1

aligned. In cases when only two representatives were compared (H6, H9, H10 and H15) the values reached at least 90% and 95%, respectively.

The complete amino acid sequence alignment of avian representatives (Table 1) of all 15 subtypes is shown in Fig. 1. HA is formed by two chains or subunits, HA1 and HA2 [7,22,23] connected by a disulphide bridge (Cys4–Cys468 in terms of H1 numbering). The two subunits result from a proteolytic cleavage activating the synthesised HA [10]. In general, the sequences of the HA2 subunit are better conserved, whereas those of HA1 exhibit higher variability (Fig. 1). This is due to the roles the two subunits play in the HA molecule (Fig. 2). The HA2 chain anchors the whole structure in the virus membrane. The higher sequence variability of HA1 reflects the evolutionary pressure since this subunit is responsible for the immune response of the host.

Table 2. Sequence characteristics of the individual subtypes of HAs

Subtype	Consensus	Identity (%)	Similarity (%)
H1 (52)	550	50	66
H2 (26)	547	76	86
H3 (39)	540	67	78
H4 (12)	542	84	90
H5 (29)	552	72	81
H6 (2)	550	91	95
H7 (17)	552	64	78
H9 (2)	544	90	95
H10 (2)	544	93	95
H13 (4)	548	85	93
H15 (2)	552	96	98

The number in parenthesis indicates the number of sequences of a given subtype used for the comparison. For subtypes that are not shown in the table (H8, H11, H12 and H14), only one representative was used.

The active site cleft of HA is formed by a receptor-binding cavity with its right edge and left edge (131_GVTAA and 221_RGQAGR, respectively; H1 numbering). Amino acid residues within the active site are highly conserved among the different HA subtypes [22]. However, in the frame of all 191 HA sequences compared in this study only the glycine is invariantly conserved in the right edge region (Table 3). The consensus sequence for both edges was found to be Gvssa and rgqsgr, respectively. It should be pointed out, however, that the arginine positioned at the N-terminus of the left edge is substituted in more than 80 cases by asparagine, and all the three serine residues alternate frequently with threonine or alanine (Table 3). With regard to the seven residues interacting with the host cell receptor (Tyr91, Trp150, Thr152, His180, Glu187, Leu191 and Tyr192; H1 numbering) only the position occupied by Thr152 in H1 exhibits low conservation. Interestingly, the threonine residue found in almost one half of all HA sequences studied is in approximately second-half substituted by a hydrophobic residue (valine, leucine and isoleucine). There are subtypes with either threonine eventually histidine (H2, H3) or a hydrophobic residue (H5, H7) as well as there are those containing both types of amino acid (H1, H4). The rest six residues from the receptor-binding site are conserved in more than 75% of HAs (Table 3).

In order to draw the evolutionary relationships among the all 15 subtypes of the influenza virus A, the evolutionary trees were calculated based on the alignment of HA sequences of the avian representatives. Three trees were constructed (Fig. 3): the tree

Table 3. Comparison of amino acid residues around the receptor-binding site

	Amino acid number						
Right edge	131 G (191)	132 V (80)	133 S (112) T (79)	134 S (65) A (45)	135 A (166)		
Left edge	221 R (101) N (82)	222 G (178)	223 Q (151)	224 S (96) A (48)	225 G (157)	226 R (186)	
Receptor binding	91 Y (190)	150 W (191)	152 T (84) V/L/I (87)	180 H (191)	187 E (148)	191 L (185)	192 Y (191)

The amino acid number denotes the position in the sequence of H1 subtype. The numbers in parenthesis indicate the number of occurrences of a given amino acid at conservative position.

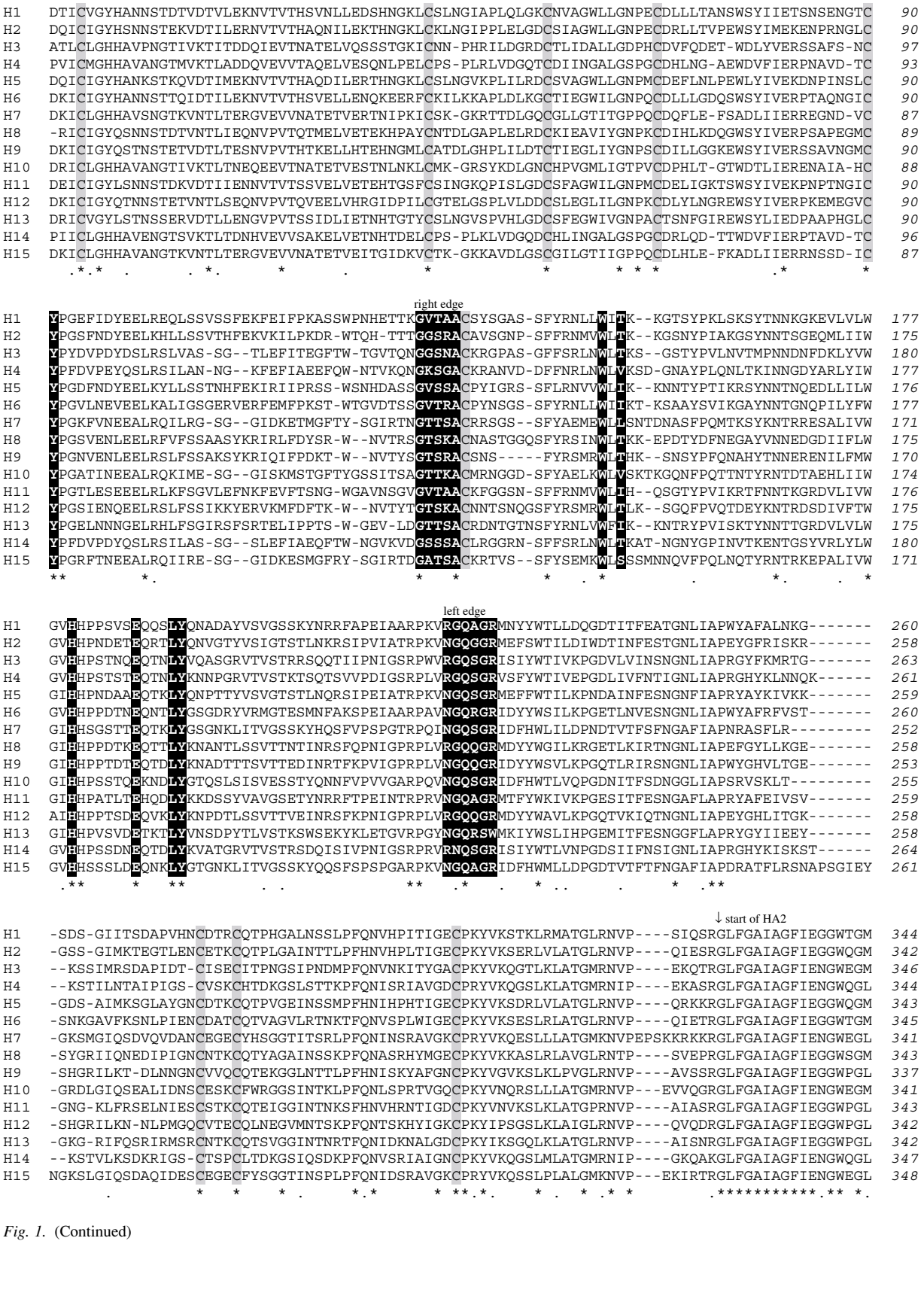

Fig. 1. (Continued)

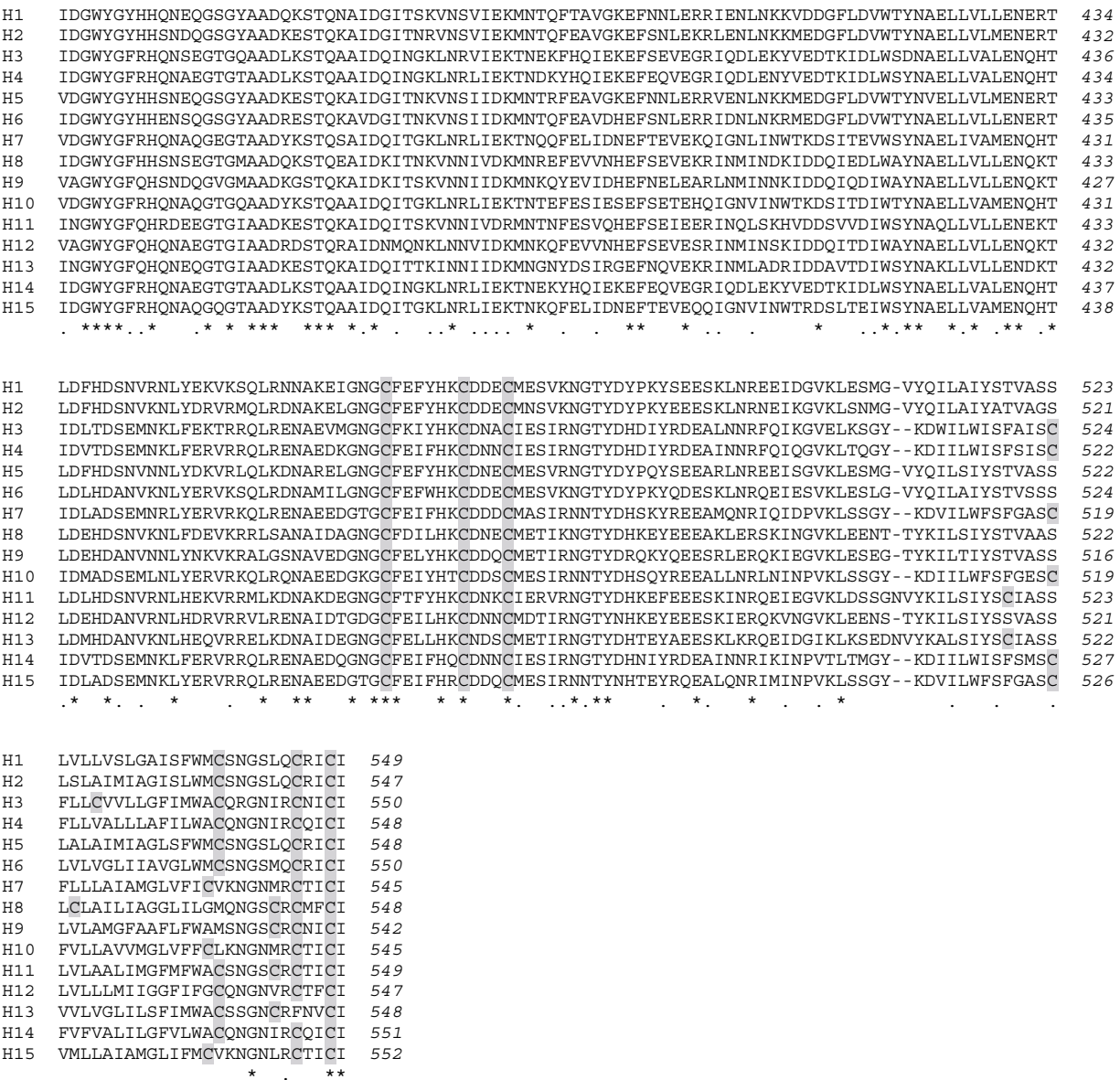

Fig. 1. Amino acid sequence alignment of 15 avian HA representatives. The sources of the individual avian HAs are explained in Table 1. The residues belonging to the receptor-binding site as well as to the right and left edges are highlighted in black-and-white inversion. The cysteine residues are highlighted in grey. The start of the HA2 subunit is indicated by vertical arrow. Asterisks and dots denote the identical residues and conservative substitutions, respectively.

based on the alignment of the entire HA sequences (i.e. both subunits HA1 and HA2) and the two partial trees reflecting the evolutionary relationships among the HA1 and HA2 subunits independently. The main observation is that the clustering and branching is very similar in all the trees. In fact, the only 'moving' subtype is the H9 subtype which in the entire HA tree is found on the branch next to the H8–H12 node whereas in the HA2 tree H9 subtype is located on an isolated branch quite away from the H8–H12 node. The situation for H9 subtype in the HA1 tree is very close to the one in the entire HA tree so that H9 is on the common branch with H12 adjacent to the H8 branch, i.e., in both HA and HA1 trees the three

subtypes H8, H9 and H12 are clustered together. Such a remarkable behaviour of the H9 subtype could be explained by a possibility that, during its evolution, the two H9 subunits were derived from two different influenza virus A subtypes. But, because in recent influenza A viruses HA1 and HA2 are coded by one and the same RNA segment, this would presume a fusion, early in evolution, of two separated RNA segments, each coding for HA1 and HA2 subunit, respectively, or a rare intrasegment recombination [24]. The rest of the subtypes exhibit very comparable evolutionary behaviour in the entire HA tree as well as in the partial HA1 and HA2 trees (Fig. 3). It is worth mentioning, however, that the two evolutionary 'groups' of subtypes can be seen in the entire HA tree and, more clearly, in the partial HA2 tree: H3, H4, H7, H10, H14 and H15, and the rest. This division is not so easily identifiable in the partial HA1 tree.

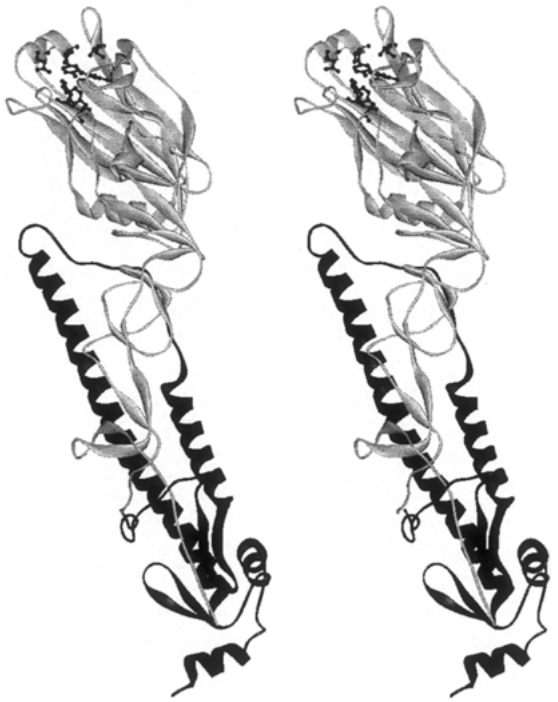

Fig. 2. The three-dimensional ribbon diagram of the influenza virus A HA. The two subunits, HA1 and HA2, are coloured grey and black, respectively. The side-chains of residues involved in the receptor-binding site (Tyr91, Trp150, Thr152, His180, Glu187, Leu191 and Tyr192) are shown in black. Data retrieved from the Protein Data Bank [28], code: 2VIU [23]. Figure in stereo prepared using the program WebLab ViewerLite 4.0, Molecular Simulations, Inc.

In an effort to trace the eventual source of the infection in mammals (i.e. also in humans), we focused on the individual subtypes with at least ten different sequences available (Table 2). Six evolutionary trees were constructed (not shown) for subtypes H1–H5 and H7 illustrating the relatedness among the HA sequences from human (mammalian) viruses and their avian counterparts. Table 4 shows the most relevant examples. Thus for instance, for the subtype H1 responsible for the 'Spanish' influenza in 1918 [3,25] several duck-originated strains should be taken into account as possible primordial natural host of the human virus. With regard to the subtype H5 that caused the death of six people in Hong Kong in 1997 [26], two chicken strains seem to be evidently responsible for the transmitting the infection. For the subtype H7 (Table 4), a possibility of direct natural transmission of an avian virus from turkey to human in Ireland was reported [27].

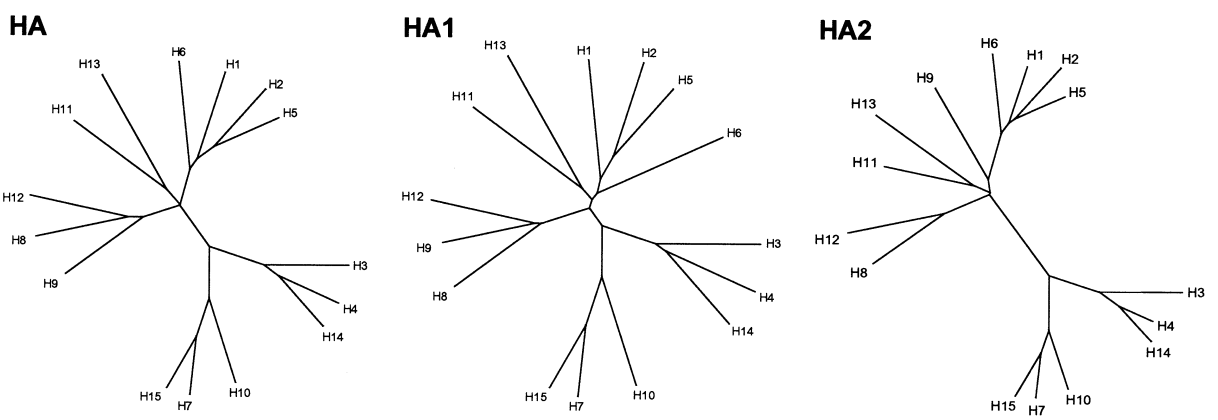

Fig. 3. The evolutionary trees of 15 avian hemagglutinin representatives. The trees were based on alignment of the entire sequences of hemagglutinins (HA), the sequences of HA1 subunit (HA1) and those of HA2 subunit (HA2).

Table 4. Human or mammalian HAs and their avian relatives

Subtype	Human/mammalian HA (GenBank)	Avian closest relatives (GenBank)
H1	A/SouthCarolina/1/18/ (AF117241)	Duck (D00839, L25071, L25072, AF091309, AF091312, AF091313), Mallard (AF091311), Turkey (AF091310)
H3	A/seal/MA/3984/92 (L32024), A/seal/MA/3911/92 (L31949)	Duck (M73771, M73772)
H4	A/swine/Ontario/01911-1/99 (AF285885), A/swine/Ontario/01911-2/99 (AF285883)	Turkey (M25290)
H5	A/HongKong/156/97 (AF028709) A/HongKong/483/97 (AF046097)	Chicken (AF082036, AF046100) Chicken (AF082035, AF046080, AF057291, AF082034, AF046099)
H7	A/England/268/96 (AF028020)	Turkey (AF028021)

In summary, the results obtained by the analysis of amino acid sequences of 191 different isolates of all available 15 subtypes of influenza virus A have shown that: (i) the active site located in the HA1 subunit is well conserved despite the great sequence variation observed among the 191 sequences of HA1; (ii) the position of Thr152 (H1 numbering) in the receptor-binding site allows the substitution for a hydrophobic residue; (iii) the HA of subtype H9 occupied different positions in the evolutionary trees based on the alignment of individual subunits; (iv) the rest of subtypes exhibited comparable clustering in all the tree trees. In order to contribute to the whole evolutionary picture of the influenza virus A, the work on the analysis of neuraminidase sequences is under way.

Acknowledgements

This work was partially supported by the VEGA grants No. 2/6045/20 and 1/7673/20.

References

1. Suarez D.L., Vet Microbiol *74*, 15–27, 2000.
2. Zambon M.C., J Antimicrob Chemother *44(Suppl. B)*, 3–9, 1999.
3. Taubenberger J.K., Reid A.H., Krafft A.E., Bijwaard K.E., and Fanning G., Science *275*, 1793–1975, 1997.
4. Taubenberger J.K., Reid A.H., and Fanning G., Virology *274*, 241–245, 2000.
5. Lamb R.A. and Choppin P.W., Ann Rev Biochem *52*, 467–506, 1983.
6. Laver W.G. and Valentine R.C., Virology *38*, 105–119, 1969.
7. Wilson I.A., Skehel J.J., and Wiley D.C., Nature *289*, 366–373, 1981.
8. Huang R.T., Rott R., and Klenk H.D., Virology *110*, 243–247, 1981.
9. Lenard J. and Miller D.K., Virology *110*, 479–482, 1981.
10. Lazarowitz S.G., Golberg A.R., and Choppin P.W., Virology *56*, 172–180, 1973.
11. Rogers G.N. and Paulson J.C., Virology *127*, 361–373, 1983.
12. Nobusawa E. and Nakajima K., Virology *167*, 8–14, 1988.
13. Air G.M., Proc Natl Acad Sci USA *78*, 7639–7643, 1981.
14. Kawaoka Y., Yamnikova S., Chambers T.M., Lvov D.K., and Webster R.G., Virology *179*, 759–767, 1990.
15. Nobusawa E., Aoyama T., Kato H., Suzuki Y., Tateno Y., and Nakajima K., Virology *182*, 475–485, 1991.
16. Röhm C., Zhou N., Suss J., Mackenzie J., and Webster R.G., Virology *217*, 508–516, 1996.
17. Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., Rapp B.A., and Wheeler D.L., Nucleic Acids Res *28*, 15–18, 2000.
18. Thompson J.D., Higgins D.G., and Gibson T.J., Nucleic Acids Res *22*, 4673–4680, 1994.
19. Saitou N. and Nei M., Mol Biol Evol *4*, 406–425, 1987.
20. Felsenstein J., Evolution *39*, 783–791, 1985.
21. Page R.D., Comput Appl Biosci *12*, 357–358, 1996.
22. Weis W., Brown J.H., Cusack S., Paulson J.C., Skehel J.J., and Wiley D.C., Nature *333*, 426–431, 1988.
23. Fleury D., Wharton S.A., Skehel J.J., Knossow M., and Bizebard T., Nature Struct Biol *5*, 119–123, 1998.
24. Rohde W. and Scholtissek C., Arch Virol *64*, 213–223, 1980.
25. Reid A.H., Fanning T.G., Janczewski T.A., and Taubenberger J.K., Proc Natl Acad Sci USA *97*, 6785–6790, 2000.
26. Lin Y.P., Shaw M., Gregory V., Cameron K., Lim W., Klimov A., Subbarao K., Guan Y., Krauss S., Shortridge K., Webster R., Cox N., and Hay A., Proc Natl Acad Sci USA *97*, 9654–9658, 2000.
27. Banks J., Speidel E., and Alexander D.J., Arch Virol *143*, 781–787, 1998.
28. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., and Bourne P.E., Nucleic Acids Res *28*, 235–242, 2000