OXFORD

Briefings in Functional Genomics, 17(6), 2018, 415–427

doi: 10.1093/bfgp/elx042
Advance Access Publication Date: 13 December 2017
Review paper

The structural variability of the influenza A hemagglutinin receptor-binding site

Michal Lazniewski, Wayne K. Dawson, Teresa Szczepińska, and Dariusz Plewczynski

Corresponding authors. Michal Lazniewski, Laboratory of Functional and Structural Genomics, Centre of New Technologies, University of Warsaw, Banacha 2c, Warsaw 02-097, Poland. E-mail: m.lazniewski@cent.uw.edu.pl, Dariusz Plewczynski, Laboratory of Functional and Structural Genomics, Centre of New Technologies, University of Warsaw, Banacha 2c, Warsaw 02-097, Poland. E-mail: dariuszplewczynski@cent.uw.edu.pl

Abstract

Hemagglutinin (HA) is a transmembrane protein of the influenza A virus and a key component in its life cycle. The protein allows the virus to enter a host cell by recognizing specific glycans attached to transmembrane proteins of the host, which leads to viral endocytosis. In recent years, significant progress has been made in understanding the structural relationship between changes in the HA receptor-binding site (RBS) and the sialylated glycans that bind them. Several mutations were identified in the HA RBS that allows the virus to change host tropism. Their impact on binding the analogs of human and avian receptors was determined with X-ray crystallography. In this article, we provide a short overview of the HA protein structure and briefly discuss the adaptive mutations introduced to different HA subtypes.

Key words: influenza A virus; hemagglutinin; host tropism; adaptive mutations

Introduction

The influenza A virus (IAV) is a single-stranded negative-sense RNA virus that belongs to the Orthomyxoviridae family [1]. The virus consists of a host-derived lipid bilayer envelope and a nucleocapsid consisting of an inner shell of matrix proteins and the viral genome that is divided into eight segments. Several proteins are encoded in viral genome including three transmembrane envelope proteins: hemagglutinin (HA), neuraminidase (NA) and the proton channel protein M2. The HA and NA proteins constitute the main antigens of IAV [2]. Up to 2001, seven other proteins had been identified to be encoded in viral genome: the nucleoprotein (NP), three subunits of viral polymerase (PA, PB1, PB2), matrix protein 1 (M1), nonstructural protein 1 (NS1) and nuclear export

Michal Lazniewski is a researcher at the Laboratory of Functional and Structural Genomics at the University of Warsaw in the Center of New Technologies (CeNT) and Department of Physical Chemistry in the Faculty of Pharmacy at the Medical University of Warsaw. His research is in bioinformatics and structural biology with a focus on medical applications using bioinformatics.
Wayne K. Dawson is a contributing researcher at the Bio-information Lab in Yayoi campus at the University of Tokyo and also associated with Center of New Technologies (CeNT) in Warsaw, Poland. His research is primarily directed to RNA structure prediction and structural biology and recently focused on chromatin structure.
Teresa Szczepińska is a bioinformatician. She has defended her PhD thesis at Nencki Institute of Experimental Biology, Warsaw, Poland in 2012. She was a postdoc in Professor Andrzej Dziembowski Laboratory at Institute of Biochemistry and Biophysics, Warsaw and now is a postdoc in Professor Dariusz Plewczyński Laboratory at Center of New Technologies, Warsaw. Her main interest is genomic data analyses.
Dariusz Plewczyński Is a professor at the University of Warsaw in the Center of New Technologies CeNT, Warsaw, Poland, and the head of the Laboratory of Functional and Structural Genomics. His main expertise covers computational genomics, biostatistics and bioinformatics. He is actively developing computational intelligence algorithms, performing biophysical simulations and applying computational modeling to various interdisciplinary problems in human genomics. He received an MS degree in Theoretical Physics (Department of Physics, UW) in 1995. In 2001, he defended his PhD degree in Physical Chemistry (Institute of Physical Chemistry, Polish Academy of Sciences), and he received a DSc degree (habilitation) in Computer Science and bioinformatics in 2012 at the Institute of Computer Science Polish Academy of Sciences.

© The Author(s) 2017. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

protein (originally called NS2). Since 2001, however, eight new proteins in particular strains have been identified, including PB1-F2, PB1-N40, PA-X, NS3, PA-N155, PA-N182, M42 and most recently PB2-S1 [3]. The functions of these novel proteins in influenza virus replication as well as pathogenesis have not been fully established. Among the viral proteins PB2-S1, M2, M42, NS2 and NS3 were detected as products of alternative splicing of mRNA, PB1-F2, PB1-N40, PA-N155 and PA-N182 are expressed via an alternative translation initiation process and PA-X is expressed via a ribosomal frameshift [4].

The influenza virus remains one of the key threats to public health. It is responsible for seasonal epidemics, which in United States alone result in 3000 up to 49 000 deaths a year, and according to some estimates nearly half a million worldwide [5]. Beside seasonal epidemics, the virus gives rise to pandemics, with serious implication for human health. Flu pandemics are believed to result from a situation where a strain that normally circulates within an avian population, for example domestic poultry, acquires the ability to efficiently infect humans via airborne transmission [6]. This is predominantly the consequence of shifting HA preference to human-type sialosides [7, 8]. Such a situation occurred only three times in 20th century. In 1918, the 'Spanish flu' pandemic claimed over 20 million casualties [9] and was caused by the H1N1 virus, where H1 indicates subtype number 1 of HA and similarly N1, subtype 1 of NA. Since that time, the H1N1 viruses have been responsible for recurrent seasonal epidemics, although the virus disappeared from the human population around 1957 and reappeared in 1977, causing the Russian flu epidemics. The 1957 'Asian flu' was caused by the H2N2 strain, which appeared to be the reassortant virus of a previously circulating H1N1 strain, with an avian virus that provided RNA segments encoding the HA, NA and PB1 proteins [10]. H2N2 itself underwent a reassortment with another avian virus, which provided fragments encoding the HA and PB1 proteins resulting in the formation of H3N2 responsible for the 'Hong Kong flu' in 1968 [10]. Since then, H2N2 has not been detected in humans, while H3N2 co-circulates with human H1N1, but further evolution of the H3N2 subtype in recent years has resulted in significantly decreased avidity [11]. Both pandemics ended in tens of thousands of deaths in the United States alone [12]. In 2009, a pandemic of 'swine flu' (H1N1) resulted in 18 449 deaths globally [13]. The H1N1 from 2009 is a result of several reassortment events between human seasonal H3N2 with avian- and swine-originating H1N1, which have been circulating in North America and Europe for decades [14, 15].

Strains carrying the abovementioned subtypes of HA and NA are responsible for the majority of infections, as only the H1N1, H2N2 and H3N2 viruses gained the ability to transmit between humans. Although the IAV strains circulating among avian species are not capable of being transmitted between humans, zoonotic infections are possible as a consequence of direct contact with an infected animal. Such rare events often lead to patient death, despite antiviral treatment because humans generally lack immunity to avian viruses. Since 1997, the most prevalent infections have been because of H5N1 and H7N9 subtypes. For example just between the beginning of 2014 to the April of 2015, 191 zoonotic infections of H5N1 were recorded and since March 2013 662 cases of H7N9 infections. The mortality rates were 32 and 31%, respectively (http://www.who.int/influenza/human_animal_interface/en). Other zoonotic infection include (H6N1) 1 case in 2013 of a Taiwanese women who eventually recovered [16]; (H5N6) 3 cases in China were recorded in 2014 with 2 fatalities; (H7N7) 89 cases in The Netherlands with 1 death followed by 3 cases in Italy in 2013 [17, 18]; (H9N2) was first recorded in South China in 1998 [19] with 7 new cases reported in both China and Egypt since 2013; and finally (H10N8) 3 cases reported until today.

Genetic drift and genetic shift are the major concerns in flu treatment. The former leads to substitution of the amino acids on the surface of the viral transmembrane proteins (HA, NA and M2) that are recognized by existing antibodies. The latter is a consequence of a process known as reassortment; a host cell can be infected by more than one strain of the virus simultaneously, on which genome segments from different strains may be exchanged as though in a mixing vessel [20]. Vaccination of the human population remains a key component in protecting against IAV [5, 21–24]. However, identifying epitopes of IAV that will circulate among humans in the following years is an extremely complicated task. A second line of defense is with antiviral drugs. Four compounds were approved by the FDA for flu treatment. The first two, oseltamivir and zanamivir, are NA inhibitors, and the other two are amantadine and rimantadine, which inhibit the M2 proton channel. The topic of antiviral treatment has been a subject of recent reviews [5, 24] and is beyond the scope of this article.

IAV strains are constantly checked for their preference to a plethora of glycans in search of warning signs that a new threat is emerging. A considerable effort is also placed on explaining at the molecular level what changes in the HA receptor-binding site (RBS) are required for avian strains to gain a preference for human receptors. Despite the urgency, it is currently impossible to reliably predict the emergence of a new pandemic; new tools are needed for scientists and policymakers to evaluate the pandemic risk posed by zoonotic viruses [10]. Although various approaches based on machine learning have been proposed [25, 26], currently, we must rely on experimental techniques. In this article, we provide an overview of mutations in HA that were confirmed to change the IAV host tropism and which influence on binding of human and avian receptor analogs was verified using X-ray crystallography. We also summarize which IAV strains have their HA structure available in Protein Data Bank (PDB), inform of recently added structures and discuss how the available structural knowledge meshes with the number of known HA and data from glycan microarray experiments.

General overview of HA structure

HA is a homotrimer, where each monomer comprised two subunits HA1 and HA2, a result of proteolytic cleavage of the HA0 precursor [27] (Figure 1). The main role of the HA1 subunit is to attach the IAV to the host cell by recognizing specific glycans, on which the IAV can enter the host cells by hitchhiking on the endocytosis process. The role of HA2 is to anchor HA to the viral membrane and second to facilitate membrane fusion of the virus with the endosomal lipid bilayer. Once the virus enters an endosome, the fusion peptide, located at the N-terminus of HA2, interacts with the target lipid bilayer. Next, as a consequence of lowering the pH value to 5 in the endosome, HA adopts a low pH conformation [28, 29] allowing the endosomal and viral membranes to merge. The fusion peptide itself adopts a helical-hairpin conformation if the peptide fragment comprised at least 23 residues or a boomerang-like conformation for shorter fragments. The 20 amino acid fragments (HAfp20) has fusogenic activity, albeit weaker when compared with the 23 amino acids fragment (HAfp23) [30]. The mechanistic explanation how the fusion occurs is still an ongoing research topic with factors like the exact orientation of the fusion peptide with respect to the endosome lipid bilayer and the degree of its penetration not completely resolved. Investigations on the truncated HAfp20 rather support an orientation nearly parallel to the surface of the lipid

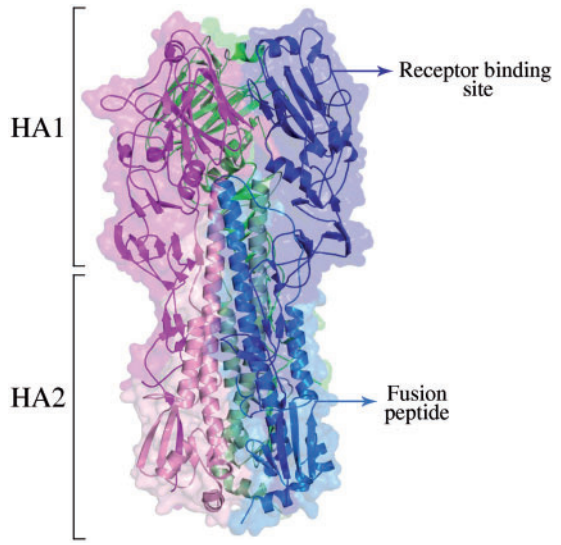

Figure 1. Trimeric structure of IAV HA, with each monomer depicted in blue, magenta and green. Each monomer is composed of the globular HA1 domain and the stalk-like HA2 domain. The location of RBS (in HA1) and fusion peptide (in HA2) are marked.

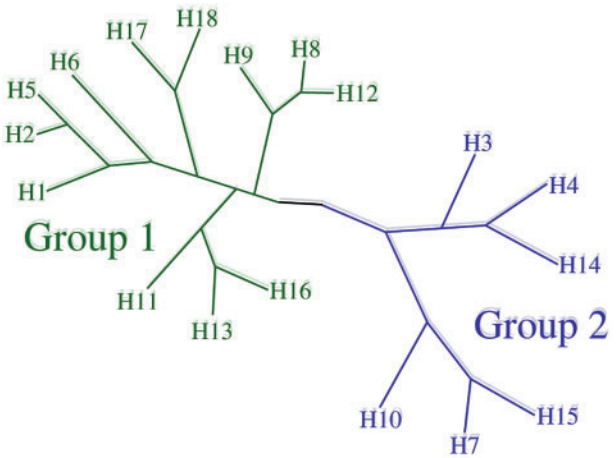

Figure 2. Schematic representation of IAV HA phylogenetic tree. The HAs are divided into two distinct groups with HAs from both groups capable of acquiring the ability to switch preference from avian-type to human-type receptors. Each group is further divided to smaller subgroups.

layer [30, 31]. However, it has been recently shown by means of molecular dynamics simulations and spectroscopic measurements that HAfp23 adopts an orientation perpendicular to the membrane plane with deeply inserted peptides [32].

To date, 18 antigenic subtypes of HA were discovered: numbered 1–18. Based on phylogenetic analysis, these subtypes were divided into two groups (Figure 2). Group 1 encompasses H1, H2, H5, H6, H8, H9, H11, H12, H13, H16, H17 and H18 subtypes, while Group 2 contains H3, H4, H7, H10, H14 and H15 [33]. Subtypes 1–16 were identified in avian species, while H17 and H18 are exclusive to bats [2]. Some IAV strains can also infect other species, such as dogs, horses, swine and humans. Identification of bat-derived IAV can potentially modify our view on how IAV evolves and migrates between different geographical localizations; however, there is no evidence that the viral genomes identified in bats can be assembled into a real virus or is able to exchange genetic material with IAV from other subtypes. Furthermore, H17 and H18 lack the ability to bind glycans terminated with sialic acids; thus, it was recently postulated that H17 and H18 should be considered to be only HA-like proteins [34].

RBSs for all 16 active HA subtypes are localized in the globular domain of the HA1 subunit. The RBS is relatively shallow and is built from a so-called 'floor', i.e. four amino acids that are conserved between all but the H17 and H18 subtypes: Y98, W153, H183 and Y195 (amino acid numbering in this work is from H3 unless otherwise stated). These amino acids are surrounded by four structural elements, namely, 130-loop, 150-loop, 190-helix and the 220-loop [7, 35] (see Figure 3A and B for details). Although all these elements are always present, their length and amino acid composition vary between strains and are often a key determinant of which type of receptors will be preferentially recognized.

Since the first crystal structure of HA of influenza virus A was solved in 1981 for HA from the A/Hong Kong/1968(H3N2) strain [36], >350 unique entries are now available in the PDB. Among them, 307 structures have their HA1 and HA2 subdomains solved with the electron density available for most polypeptide chain. However, these structures represent only 73 strains, which is only a fraction of all the available HA sequences (summarized in Table 1). The H4, H8, H11 and H12 subtypes have no available three-dimensional (3D) structures in the PDB. The HA structures (even from different subtypes) are relatively similar, for example the root mean square deviation (RMSD) for HAs from H2 and H3, which belong to different phylogenetic groups, is only 2.4 Å. All the available HA structures were solved with X-ray diffraction, and overall they are of good quality. For >85% of the structures, the resolution is better than 3 Å and, for nearly 88.5% of them, the *R*-free value is <0.3.

In past 2 years, several new structures of HA were solved. Specifically, in 2017, the structure of HA from the H15 subtype was solved in complex with the avian receptor analog 3′SLN (NeuAcα2-3Galβ1-4GlcNAc) and in the apo form. In 2016, the structure of a potential drug arbitol in complex with HAs from the H7N9 and H3N2 viruses was proposed and compared with structures solved with another fusion inhibitor TBHQ. In that work, a pocket between the monomers of the HA trimer was described, which can be targeted by other small compounds to stabilize the HA2 in its normal pH conformation. Additionally, a complex of several HAs with an antibody capable of neutralizing diverse subtypes of IAV was solved. The structure of HA from the H10 subtype with a preference toward human-type receptors was proposed and compared with a structure from the wild-type H10. The role of the 150-loop in modifying preference for HAs belonging to the H7, H10 and H15 subtypes was highlighted.

For easier access to PDB resources, several influenza- or HA-specific databases are available. Among them is the Influenza Research Database (IRD) [37], which provides access to several tools allowing sequence and structural analysis of all proteins from IAV. 3DFlu [38] is a database that allows the user to search the PDB for HAs from specific strains or using parameters, such as localization, infected host or year of infection. It also offers interactive graphs for analysis of the relation between HAs, based on their sequence, structural or electrostatic potential. HASP (Hemagglutinin Structure Prediction Server) [39] is a collection of 3D models for all known HA protein sequences, built using Rosetta. It allows the user to compare and visualize associated models.

Role of sialylated glycans in IAV infection

Mammalian cells are covered by a glycocalyx, which comprises glycolipids, glycoproteins, glycophospholipid anchors and

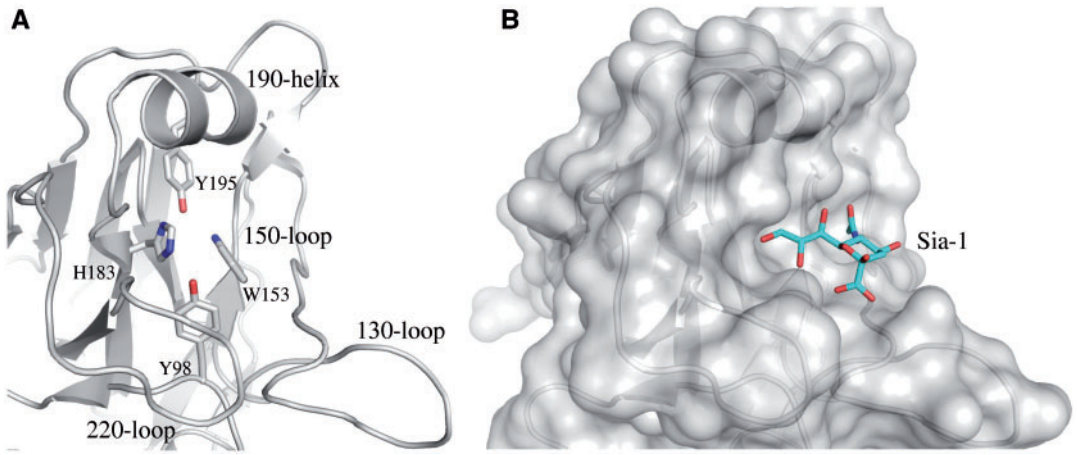

Figure 3. The structure of HA RBS. (**A**) Four amino acids conserved between HAs of different subtypes are marked with their numbering according to H3. The secondary structure element surrounding the binding site is also labeled. (B) The same RBS represented as a surface. It can be seen that the RBS is relatively shallow. The typical location of Sia-1 in the RBS is shown.

Table 1. Summary of all strains for which 3D structure of full-length HA is available in the PDB

Subtype	Strain	Number of entries in PDB
H1N1	A/South Carolina/1/1918	6
	A/Puerto Rico/8/1934	3
	A/swine/Iowa/1930	2
	A/duck/Alberta/1976	1
	A/California/04/2009	13
	A/WDK/JX/12416/2005	4
	A/Darwin/2001/2009,	1
	A/Solomon Islands/3/2006	1
	A/Korea/01/2009	1
	A/Thailand/CU44/2006	1
	A/Washington/5/2011	1
	Total	42
H2N2	A/Chicken/New York/29878/91	3
	A/chicken/Potsdam/4705/1984	1
	A/Japan/305/1957	12
	A/Singapore/1/1957	3
	Total	19
H2N3	A/swine/Missouri/2124514/2006	1
	Total	1
H3N8	A/Canine/Colorado/17864/2006	7
	A/duck/Ukraine/1963	3
	A/Equine/New market/2/1993	4
	A/Equine/Richmond/1/2007	4
	A/harbor seal/Massachusetts/1/2011	2
	Total	20
H5N1	A/Anhui/1/2005	1
	A/chicken/Hong Kong/YU562/2001	1
	A/chicken/Vietnam/NCVD-093/2008	3
	A/common magpie/Hong Kong/5052/2007	1
	A/duck/Egypt/10185SS/2010	3
	A/duck/Guangxi/2396/2004	1
	A/duck/Laos/3295/2006	1
	A/Duck/Singapore/3/1997	3
	A/Egypt/N03072/2010	1
	A/Goose/Guangdong/1/1996	2
	A/goose/Guiyang/337/2006	1
	A/goose/Hong Kong/437-10/1999	1
	A/Hubei/1/2010	1
	A/Indonesia/5/2005	6
	A/turkey/Turkey/1/2005	7
	A/Vietnam/1203/2004	12
	A/Vietnam/1194/2004	22
	A/Xinjiang/1/2006	1
	Total	68
H5N8	A/gyrfalcon/Washington/41088-6/2014	1
	Total	1
H6N1	A/chicken/Taiwan/A2837/2013	3
	A/Taiwan/1/2013	3
	A/Taiwan/2/2013	10
	Total	16
H6N2	A/chicken/New York/14677-13/1998	2
	Total	2
H6N6	A/Chicken/Guangdong/S1311/2010	2
	Total	2
H7N2	A/New York/107/2003	4
	Total	4
H7N3	A/turkey/Italy/2002	1
	A/turkey/Italy/214845/2002	4
	Total	5
H7N7	A/Netherlands/219/2003	4
	Total	4
H7N9	A/Anhui/1/2013	13
	A/Shanghai/2/2013	11
	A/Shanghai/1/2013	5
	Total	29
H9N2	A/Swine/Hong Kong/9/1998	3
	Total	3
H10N2	A/mallard/Sweden/51/2002	4
	Total	4
H10N7	A/green-winged teal/Texas/Y171/2006	1
	Total	1
H10N8	A/Jiangxi/IPB13a/2013	7
	A/Jiangxi-Donghu/346/2013	8
	Total	15
H13N6	A/gull/Maryland/704/1977	2

(Continued)

Table 1. (Continued)

Subtype	Strain	Number of entries in PDB
	Total	2
H14N5	A/mallard/Astrakhan/263/1982	2
	Total	2
H15N9	A/shearwater/Western Australia/2576/1979	2
	Total	2
H16N3	A/Black-Headed Gull/Sweden/2/99	2
	Total	2
H17N10	A/little yellow-shouldered bat/Guatemala/060/2010	2
	Total	2
H18N11	A/flat-faced bat/Peru/033/2010	2
	Total	2

Note: The values represent the number of unique entries associated with a given strain.

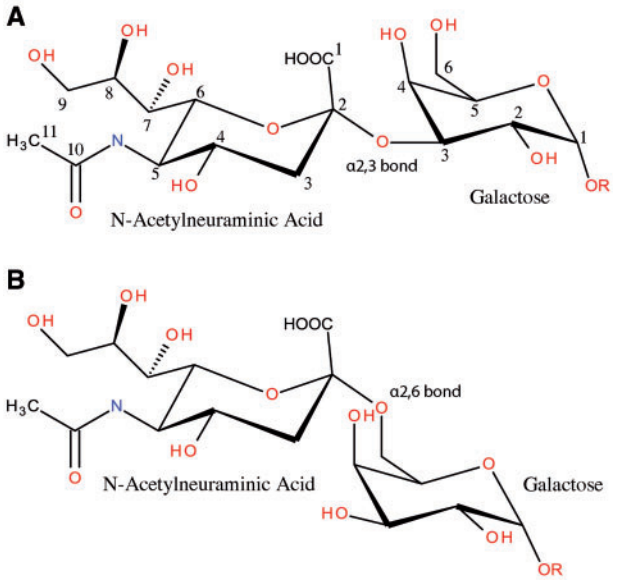

Figure 4. The ligands that are recognized by HA. (**A**) Avian-type receptors, which are preferentially bound by avian-infecting IAV, are terminated with sialic acid (usually N-acetylneuraminic acid, Sia-1) connected to the penultimate galactose (Gal-2) with an α2,3 bond. (**B**) For human-type receptors, recognized by human infecting IAV, Sia-1 and Gal-2 are connected with an α2,6 bond. IAV infecting swine can recognize both avian- and human-type receptors. The numbering scheme for galactose and N-acetylneuraminic acid is provided.

proteoglycans. There are several types of glycans, including N-glycans, O-glycans. Some of them are terminated with a galactose connected to sialic acids. The main sialic acid in humans is N-acetylneuraminic acid. These two components are often connected with either an α2,3 or α2,6 glycosidic bond (Figure 4). In the 1990s, it was demonstrated that the human upper respiratory tract including the trachea appears to contain mainly α2,6 receptors [40]. Human bronchial and lung tissues have both α2,6- and α2,3-linked glycans, although the lungs of children showed more α2,3 terminated glycans [41]. Both the α2,3- and α2,6-terminated glycans are expressed in swine in the respiratory tract. Avian species have α2,3- and α2,6-linked glycans in both the respiratory and intestinal tract, although differences in the abundance of these receptors between different species and tissues were reported [42]. An important component of the respiratory system is also mucin, produced by the goblet cells. Mucin is a conglomerate of glycoproteins, which have both α2,3- or α2,6-sialylated glycans [43]. It is interesting to note that although it is a well-established fact that the influenza virus recognizes the host cell by binding to sialylated glycans, IAV was also able to infect cells, which were treated with sialidase and therefore should have their sialic acid receptors removed [40, 44].

The main difference distinguishing IAV circulating in human and avian species is the preferential recognition of glycans terminated with the α2,6 linkage by the former and α2,3 linkage by the latter. Thus, the α2,6-terminated glycans are often called 'human receptors', while the 'avian receptors' name is restricted to receptors with the α2,3 linkage. IAV infecting swine can recognize both avian and human receptors. The binding of HA is still relatively weak with a K_d depending on the ligand type and tested strains ranging from the millimolar (mM) to micromolar (μm) scale [45]. In this context, it is worth considering that the topology and density of the glycans on the membrane surface as well as the time of incubation are important factors for binding HA in a successful infection.

The ability to switch the preference from avian to human receptors is considered to be a key element necessary for a virus to cause human pandemics. This event allows the virus to replicate in upper respiratory tract, avoid ineffective binding to mucin and triggers host pathways, such as the stimulation of sneezing, leading to efficient viral transmission between humans [40]. Recent systematic analysis of glycan arrays [46] also points out that human-infecting strains can also bind Neu5Acα2–8Neu5Acα2–8Neu5Ac and Neu5Gcα2–6Galβ1–4GlcNAc substructures. Moreover, most IAVs can recognize Galβ- and GlcNAcβ-terminated glycans. These observations suggest that it might be the structural pattern of the polysaccharides instead of N-acetylneuraminic acid, which is recognized by influenza viruses of various host origins. It must be noted that other factors also influence the ability of IAV to cross an interspecies barrier, among them the interplay between HA and the other transmembrane protein NA. The NA role, although typically associated with the effective release of the nascent virus particles from the surface of the infected cell [47], is in fact much broader. It additionally plays a role before the infection by trimming glycosylation of HA and allowing the virus to penetrate the mucin layer. Like HA, NA should gain the sialidase activity toward human receptors for effective transmission in humans. This fact was observed, for example, for the N2 subtype, which, gradually, between 1967 and 1972 acquired increased activity for the α2,6 glycosidic bond [48]. It is believed that the functional balance between HA and NA exists and is necessary for effective infection by IAV. For example, treating IAV with the NA inhibitor oseltamivir, resulted in escape mutants, among which some exhibit no change to the NA sequence, but rather to the HA protein, which showed a reduced affinity toward its receptor. Thus, the role of NA in the infection cycle was decreased [49]. More examples of such interplay were summarized in a recently published review article [50].

The majority of the structures are co-crystallized with a ligand—usually an analog of either avian or human receptors or with antibodies targeting a given strain. The most popular ligands are the linear sialopentasaccharides LSTa (NeuAcα2-3Galβ1-3GlcNAcβ1-3Galβ1-4Glc) and LSTc (NeuAcα2-6Galβ1-4GlcNAcβ1-3Galβ1-4Glc), which are analogs of avian and human receptors, respectively. Both compounds are isolated from human milk. Other analogs include 3′SLNLN (NeuAcα2-

3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAc), 6′SLNLN (NeuAcα2-6Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAc) or the shorter sialotrisaccharide 3′SLN and 6′SLN. These compounds are often used to assess how IAV recognizes the host cells; however, this approach has a potential drawback. As was mentioned already, the human respiratory tract is covered with a great number of diverse sialylated glycans. These glycans are usually longer and not always as linear as the typical ligands used in crystallography. The reasons for using the abovementioned compounds are their availability and quantity [51]; however, this does not justify using them exclusively. We must also consider the limits of our knowledge about which sialic acid receptors are relevant for infection *in vivo*. Even more in complexes of HAs with these analogs, the electron density is often visible only for a limited number of saccharides or only for the N-acetylneuraminic acid itself. This is often explained by the weak binding of a particular ligand to HA from an analyzed strain. Usage of a different polysaccharide ligand might result in a clearer electron density for an entire molecule, which would provide invaluable structural knowledge. Additionally, the glycan microarray experiments indicate that HAs exhibit significant variability in binding different sialylated glycans. Thus, it might be a significant oversimplification to assume that, for example, all avian receptors bind exactly the same way as LSTa. This should prompt us to test compounds, which differ from those typically used to understand what structural difference might be responsible for these phenomena.

Adaptive mutations of IAV HA

Historically, at least two adaptive mutations in RBS were necessary to switch the preference from avian- to human-type receptors (Figure 5): E190D and G225D for H1N1, and Q226L and G228S in the case of H2N2 and H3N2. However, those mutations are not universal, as will be explained in detail in the following paragraphs. Furthermore, it is unlikely that each HA subtype has only one set of mutations that will cause a switch in the host tropism from avian to human.

The first subtype able to cross the interspecies barrier was H1N1, which was responsible for the 1918 ‘Spanish flu’ pandemic [52]. Two mutations contributed to switching the host preference E190D, which was observed in all five HAs isolated from the 1918 pandemic, and G225D, observed in three cases. To analyze the impact of these mutation on the H1, the HA structures from two closely related strains were solved in 2004 [53] with analogs of human and avian receptors (LSTc and LSTa). The first strain, A/swine/Iowa/1930 (SW/30), has a preference for human-type receptors and was the first influenza virus isolated from mammals. Its RBS sequence was identical to two out of five HAs from the 1918 strain with G225 and D190. The second strain, A/Puerto Rico/8/1934 (PR/34), could bind both avian- and human-type receptors and is among the first human strains isolated in North America. Its RBS carries D225 and E190. The main differences between the H1 and other subtypes are the conformations of the 130- and 220-loops. As a consequence, one of the key amino acids in the RBS (Q226) adopts a ‘lower conformation’ than the corresponding L226 in the H3 subtype. Thus, the Gal-2 from the human receptor is tilted about 10° into the RBSs of the H1. For both LSTc and LSTa, the location of Gal-2 is around 1–2 Å closer (depending on the receptor type) to the 220-loop than in the other subtypes. The Gal-2 position of the avian receptor analog is stabilized by interacting with the side chain of K222 in both strains and, additionally, only for PR/34 by the nearby D225. Considering that both strains are able to bind human-type receptors, the role of G225D is probably restricted to enhancing the human-type receptor binding rather than enabling it in the first place [53]. As SW/30 exhibits only weak binding to the avian-type receptor, only the conformation of Sia-1 can be deduced from the electron density. On binding LSTa to HA from the PR/34 strain, the Q226 adopts an ‘upper conformation’, while for SW/30, no change in the Q226 conformation was observed on binding different ligands. This possible inability of Q226 in HA from SW/30 to adopt a higher position in the RBS seems to explain the failure of SW/30 HA to interact as effectively with the avian receptors. In the PR/34 complex with LSTa, E190 interacts through two water molecules with Q226. This network of hydrogen bonds may be necessary to position the Q226 in the binding site for interaction with the LSTa. In contrast, D190 of the SW/30 HA does not interact with either the 9′OH of Sia-1 or Q226 and is thus unable to facilitate binding of the avian receptor analog. Instead it might be possible that D190 interacts with the GlcNAc-3 of the human-type receptor. The H1N1 virus responsible for the 2009 pandemic (similar to H1N1 from 1918) has D190 and D225 [13]. Interestingly, the D225G substitution in the 2009 pandemic virus has a far more subtle effect on the receptor binding with respect to the HA of the 1918 influenza virus [54]. A possible explanation is the presence of E227 in the 2009 pandemic virus compared with A227 in the HA of the 1918 strain. E227 is part of a network of charged residues, including D225 and K222 that stabilizes the Gal-2. Thus, the absence of D225 may be compensated by the nearby E227 residue and may help maintain the overall conformation of the 220-loop and stabilize the interactions between K222 and human receptors.

Although viruses carrying the H2 subtype disappeared from the human population in 1968, they continue to circulate in both avian and swine reservoirs. This highlights the need to understand the potential of these viruses to spread and cause disease in humans [55]. The H2 subtypes are phylogenetically more related to H1 than the H3 subtypes, which explains why the RMSD between HA monomers of H1 and H2 is 1.3 Å compared with 2.4 Å between HA monomers of H2 and H3, the third human pandemic strain. Yet despite the relatively low overall sequence similarity between H2 and H3, which is around 40%, the H2 subtype shares with H3 the same adaptive mutations Q226L and G228S [56]. The structure of the avian receptor (LSTa) bound to the RBS of the avian-infecting strain A/Chicken/New York/29878/91 (NY/91) is similar to that of other subtypes with the ligand in the *trans* conformation between Sia-1 and Gal-2. Q226 forms several hydrogen bonds with glycosidic oxygen between Sia-1 and Gal-2 and with Gal-2 4′OH group. The human receptor (LSTc) is in the *cis* conformation with two water mediating interactions between the protein and Gal-2, e.g. a single water molecule links Q226 and N186 with the 4′OH of Gal-2 and the 9′OH of Sia-1 and second water links K222 and 3’OH group of Gal-2. The conformation of the human receptor bound to the NY/91 is remarkably similar to that observed in the complex with the human receptors preferring HA (A/Singapore/1/1957 (S1/57)). However, the Q226L and G228S mutations enhance the binding of the human receptor to S1/57. In NY/97, Q226 coordinates a water molecule, which is close to the position occupied by Gal-2 in the human HA receptor complex. This suggests that HA of S1/57 benefits from acquiring leucine at position 226 because human receptors can then bind without the need to displace the water molecule. Furthermore L226 creates a hydrophobic environment, which would promote localization of the C6 atom from Gal-2 in its vicinity, which otherwise would be directed to the solvent or (like in NY/97) would face the

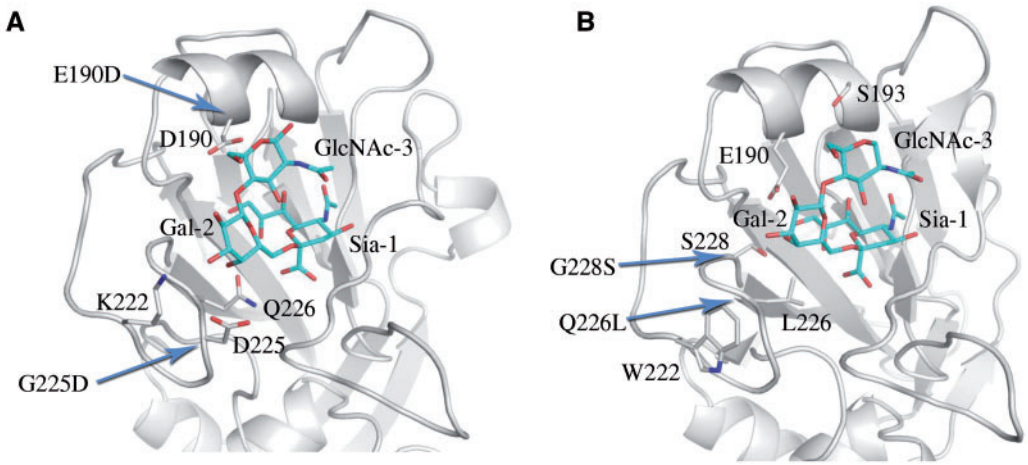

Figure 5. The HA RBSs for subtypes that were able to change their host tropism and still circulate among human population. (**A**) H1N1 represented by A/California/04/2009, (**B**) H3N2 represented by A/Aichi/2/1968. For both cases, the sialic acid (Sia-1) followed by galactose (Gal-2) and N-ectylglucose (GlcNAc-3) of LSTc are shown. In both cases, the conformation of the α2,6 linkage between Sia-1 and Gal-2 is *trans*. The residues, which role is described in the text, are depicted.

hydrophilic side chain of Q226 [56]. In the complex of S1/57 with an avian receptor analog, only the Sia-1 position could be proposed, most likely because of the poor binding of this molecule as a consequence of the Q226L mutation. It is interesting to note that the complex of NY/91 HA with LSTc is well ordered, contrary to situation observed for the avian receptor preferring strains of H1 and H3 subtypes. It was suggested to be because of the presence of N186, whether serine and proline are observed in H3 and H1, respectively. As mentioned before, N186 interacts with Q226 and the 4′OH of Gal-2 through a water molecule; it is likely that the absence of this residue and the hydrogen-bonded network that involves Q226 contributes to the relatively poor binding of the human receptor analog by the avian H1 and H3 subtypes. However, N186 does not play any major role in stabilizing the human receptor for the H2 subtype, and this amino acid adopts the same orientation as in the unligated structure. The RBS of the S1/57 subtype is also slightly elongated with respect to the NY/97 subtype; a factor that contributes to a more favorable accommodation of the human-type receptors.

Similar to the H2 subtype, Q226L and G228S allowed the HA from subtype 3 to switch its preference from avian to human receptors and to gain the ability of airborne transmission between humans. The structure of HA for the A/duck/Ukraine/1963 strain (UK/67), a possible progenitor of the pandemic virus, was solved in 2003 [57] with LSTa and LSTc and could be compared with the HA from the pandemic H3N2 strain A/Aichi/1968 (AI/68); the 3D structure was released in 1991 [58]. The RMSD difference between the HA monomers from these two strains is relatively small, a mere 0.5 Å. However, in the case of the UK/67, the binding site is slightly narrowed by a difference in the position of the main chain of the 220-loop, which is 0.5 Å closer to the 130-loop. As mentioned before, it is generally accepted that the wider RBS would promote human-type receptors because of the conformation it would adopt in the binding site. The conformation of the avian receptor analog with the avian H3 is remarkably similar to ones obtained in complexes with the avian influenza from the H2 and H5 subtypes. The ligand is in the *trans* conformation with the carboxyl group of the Sia-1, the glycosidic oxygen and the 4′OH atom of the Gal-2 pointing toward the Q226 side chain. As a consequence, the conformation of the GlcNAc-3 is also similar and points to the exit point from the RBS. The electron density for Gal-4 and GlcNAc-5 was incomplete or unobserved. The LSTc is bound to the avian H3 in the *cis* conformation with the C6 atom of Gal-2 located near polar Q226, which may account for the poor binding of the LSTc to avian HA. The human receptor analog (in complex with HA of the pandemic H3) adopts the same *cis* conformation, but the Gal-2 is located closer to the 220-loop compared with the complex with avian H3, which is most likely a consequence of the hydrophobic leucine at position 226. The electron density is visible for all the atoms of the human analog, which suggests sufficient binding of this ligand. It can be clearly seen that the ligand leaves the RBS next to the N-terminal part of the 190-helix.

In recent years, a significant decrease in avidity of strains belonging to the H3N2 lineage can be observed. The changes in the RBS affect the ability to recognize both the avian- and human-type receptors [11]. The key mutations associated with this behavior are the E190D, followed by G225D and W222R, recorded in 2001–2002, and the S193F and D225N, from 2004–2005. Simultaneously, the L226 mutated first to valine and in 2004 to isoleucine. These changes caused a 4-fold decrease in the avidity toward human receptors between strains from 1968 to 2001, followed by a 200-fold decrease in the avidity between 2001 and 2004. The strains isolated in 2010 failed to bind to human receptor under typical assay conditions. The distance between the C6 atom of Gal-2 and the side chain of residue 226 increased from 3.8 to 4.5 Å between strains from 1968 and 2004. In the 1968 human strain, E190 together with H183 interact with the 9′OH of Sia-1. In contrast, the shorter side chain of D190 in the A/Finland/486/2004 (Fi/04) HA does not reach into the pocket sufficiently to interact with the glycerol substituent of Sia-1. However, for the Fi/04, Gal-2 interacts with D225, which is stabilized by R222. To adopt this conformation, a rearrangement of the 220-loop with respect to the apo form is required; thus, it was concluded that the 220-loop of the 1968 HA is already positioned for receptor binding, while the need to rearrange the 220-loop decreases the ability to efficiently bind the human receptor in Fi/04. Further decrease in avidity was observed for strains isolated in 2005, which carried the D225N mutation. Although glutamine at position 225 can still form a hydrogen bond with R222, it does not interact with Gal-2 of the human receptor analog. Thus, it was concluded that there is insufficient binding energy to facilitate the conformational change in the 220-loop.

Since the first records of zoonotic infection by IAV from the H5N1 lineage, the naturally circulating H5N1 viruses still have not gained the ability to effectively transmit between humans [59]. Furthermore, the adaptive mutation identified previously for the HAs in the H1, H2 and H3 subtypes resulted in either losing the ability to bind sialylated glycans by H5 or showed a small increase in the ability to bind human-type receptors [60, 61]. This, however, does not exclude the possibility that the H5N1 virus could gain a preference toward human receptors by means other than those observed for the H1, H2 and H3 subtypes. In 2012, several articles were published describing the introduction of specific mutations to H5 that could result in switching the HA preference to α2,6-linked glycans. Replacing the H3N2's HA with the mutated H5 resulted in a reassortant virus that was able to transmit between ferrets, which are used as a model to assess the ability of IAV to transmit via airborne transmission. Mutated H5s, however, bind about 5-fold weaker to human receptor, and about 10-fold weaker to avian receptors than the H3 from the 1968 pandemic, virus A/Aichi/2/1968 [62]. The proposed mutations were N158D, N244K, Q266L and T318I for the A/Vietnam/1194/2004 (Vt/1194) strain [63]; H110Y, T160A, Q226L and G226S for A/Indonesia/5/2005 (Ind/05) [64]; and Q196R, Q226L and G228S for A/Vietnam/1203/2004 (Vt/1203) [65]. The effect of the first two mutations with respect to the wild-type H5 was explained on a molecular level in several recently published articles [62, 66].

The main difference between avian receptor preferring H5 (Vt/1194) and its transmissible mutants is the position of the Gal-2 of 6′SLN in the RBS. For the transmissible mutant, it closely resembles the conformation observed in the H1, H2 and H3 human pandemic strains, where the ligand adopts a *cis* conformation and exits the RBS close to the 190-helix. In a complex of 6′SLN with the wild-type H5, the ligand also adopts a *cis* conformation, but the Gal-2 is rotated ~90° about C6-C5, causing the ligand to exit the RBS next to the 130-loop. This unique Gal-2 conformation allows the formation of the bond between Gal-2 3′OH and the main chain atom of amino acid 225, which is facilitated by the low position in the site adopted by Q226. The conformation of Q226 is stabilized by a hydrogen bond between its side chain and S137. Introducing Leu at position 226 obviously disrupts these interactions and, furthermore, widens the RBS by ~1 Å. Owing to the Q226L mutation, the transmissible mutant binds the avian analog in a *cis* rather than a *trans* conformation, observed for the avian receptor preferring H5. As a consequence, the avian analog leaves the RBS closer to the 130- and 220-loops than in the wild-type H5 subtype. Three other mutations are observed in the transmissible mutant. N158D is responsible for losing the glycosylation site in the vicinity of the N-terminus of the 190-helix. The large carbohydrate chain attached to the N158 would most likely hinder the binding of the human receptor, which leaves the binding site next this location for a transmissible mutant. The side chain of the amino acid located at 224 is directed outside the RBS and does not interact directly with the human- or avian-type receptor. However, introduction of a positive charge by the N244K mutation might increase the nonspecific interactions between the virus and host cell. T318I also does not interact with the HA ligand, as it is located nearly 70 Å away from the RBS. The role of the T318I mutation is thus assumed to increase the thermostability of the protein.

In 2013, it was proposed that four key elements in the H5 RBS are required for H5 to effectively switch its preference. These include (1) modification of the 130-loop length, (2) mutation of at least one amino acid on the 130- or 220-loop, (3) at least one mutation on 190-helix and (4) a mutation of N158 [67]. In 2015, it was observed that the HA from the strain A/duck/Egypt/10185SS/2010 already has two of the required characteristics [59]: deletion of the 133a amino acid and mutation of N158. By introducing the Q226L mutation, the strain effectively switched its binding preference to the human receptors. For the second tested strain, A/chicken/Vietnam/NCVD-093/2008 (chicken/08), four mutations were introduced: N224K, Q226L, N158D and the deletion of L133a. As a consequence of these changes, HA gained human-receptor preference, and its avidity was comparable with previous pandemic strains. Analyses of the human receptor analog (LSTc) in complex with these mutants exhibited a clear electron density for all the sugars. The ligands assume a typical *cis* conformation that is almost identical to one from the complex with the pandemic H2 virus. In the previously studied transmissible H5 subtype, the backbone oxygen of amino acid 133a was shown to be in close contact with the N-acetyl group of Sia-1 (at the distance of 3.1 Å), which suggests close van der Waals (VDW) interactions. In both mutants, the single amino acid deletion causes the backbone to move away from Sia-1, which mimics the loop conformation from H2. Interestingly, insertion of the S133a to A/duck/Egypt/10185SS/2010 with the Q226L mutation or L133a to the chicken/08 virus almost completely abrogates the receptor binding to both the avian and human receptors.

In 2013, the first human infection with a virus carrying the H6 subtype was recorded in Taiwan [16]. The strains with this subtype have been present in poultry in Taiwan since the 70s, but it seems only recently that it gained the ability to infect humans. Phylogenetic analysis of the genes from this H6N1 human isolate suggests that an avian virus A/Chicken/Taiwan/A2837/2013 (ckTW/13) was provided seven of eight genes of the A/Taiwan/2/2013 (TW2) virus, whereas the eighth segment, coding for PB2, probably originated from A/Chicken/Taiwan/0101/2012(H5N2) [68]. The receptor preference for the human-infecting H6 virus is still under debate, with some research suggesting that only the glycans terminated with the α2,3 bond are recognized [68], and others showing some preference toward the α2,6 ligands [51, 69]. Comparisons of a typical avian H6 A/chicken/Guangdong/S1311/2010(H6N6) (GD/10) with the TW2 revealed several differences in the RBS alone, including S133R, S136T, S137N, I155V, P186L, V187D, E190V, A193N and G228S. Furthermore, like the H1 and H5 and contrary to H2 or H3, H6 has a bulge in the 130-loop with the main-chain carboxyl group of the amino acid at position 133 pointing into the RBS, According to the available crystallographic data, the key amino acids responsible for switching the preference to the human receptors are at positions 137, 190 and 228, which was suggested by comparing the structures of HA from TW2 and GD/10 [51] and residues 190, 228 and 186 according to a comparison of the TW2 HA crystal structure with that from ckTW/13 [69]. Thus, a new set of mutation was introduced allowing a switch in the preference of the HA protein, despite the lack of the Q226L signature mutation that is crucial for the H2, H3 and H5 subtypes.

The 3′SLNLN binds to both ckTW/13 and TW2 in a *cis* conformation with the key interactions formed with Q226, S228. For the ckTW/13 strain, an additional hydrogen bond is formed between N137 and Gal-2. It is worth mentioning that the amino acids from the 130-loop, in all but one case of a transmissible H5 mutant, do not interact directly with the Gal-2 but rather with Sia-1. In a complex of TW2 with the avian receptor analog, Gal-2 did not form a hydrogen bond with N137, most likely because of the introduction of leucine at position 186 instead of proline, which is observed in ckTW/13. The leucine is responsible for a

more hydrophobic environment for the GlcNAc-3 and affects the overall conformation of the ligand through VDW interactions. Thus, the ligand moves away from the N137, which might account for the weaker binding of the avian receptor to the TW2 strain. Interestingly, for another avian H6 (GD/10), the N137 is replaced with serine, and S228 is substituted with glycine; thus, these amino acids do not provide the polar contacts in the complex of GD/10 HA with the avian receptor analog. The authors of a second work [51], however, observed the hydrogen bond involving N137 between the avian receptor analog and TW2. For the GD/10, the side-chain carboxyl group of E190 forms a strong hydrogen bond with the 9′OH of Sia-1 from LSTa. This, together with the presence of Q226 is responsible for LSTa moving deeper into the RBS. For LSTa in complex with the TW2, the comparatively larger N137 at the front of the RBS and the hydrophobic V190 and S228 at the back of the RBS (compared with S137, E190, and G228 in GD/10) are believed to be responsible for the higher-sitting position of LSTa in the complex with the human H6. This results in weaker binding of LSTa to this strain. The discrepancy in both works involving the HA of TW2 strain might be attributed to using different analogs of avian receptors—3′SLNLN and LSTa. This highlights the need for using different sialylated glycans to provide a reliable answer on how HAs interact with the avian and human receptors *in vivo*. The LSTc in complex with the TW2 strain adopts a *cis* conformation and is similar to that observed for H2, H3 and the transmissible H5, despite the presence of Q226 instead of the typically observed L226. It seems that the combination of N137, V190 and S228 results in a similarly positioned Sia-1 moiety for LSTc, thus likely alleviating the unfavorable interaction between Q226 and the C6 atom of Gal-2. 6'SLNLN binds to both GD/10 and TW2 in an atypical *trans* conformation, although for the former, Gal-2 is rotated about 60 degrees around the C5-C6 bond.

Since 2013, several hundred zoonotic infections with the H7N9 strains were reported, albeit with a lower mortality rate than infections associated with the H5N1 viruses. The human-infecting H7N9 virus contains HA and NA viral RNA segments of duck origin with the remaining segments from the H9N2 chicken-infecting virus [70]. Understanding the mechanism to switch preference for this virus is particularly important, considering some strains of this subtype already acquired the Q226L mutation. Recent works suggest that the H7N9 virus, since 2013, is capable of at least limited droplet transmission in ferret models [71, 72], with inefficient transmission in more stringent tests. In 2013, the crystal structures of several H7N9 strains with human and avian receptor analogs (3′SLNLN and 6′SLNLN) were solved to better understand the underlying mechanism behind the switch in receptor preference [73, 74]. Analysis of the A/Shanghai/1/2013 strain (SH-1), which has the avian-signature residue Q226, reveals that this strain has a preference toward avian receptors. The strain A/Anhui/1/2013 (AH-1) exhibits a dual binding property, and its HA carries L226. In work by Shi *et al.* [74], the structures of SH-1 were proposed and compared with that from AH-1. Their RBS differs by only four residues: S138A, G186V, T221P and Q226L. Interestingly, the Q226L mutation appeared not to be a sole determinant for the host tropism for the H7 subtype, as the L226Q mutant could still bind to human receptor analogs, although with reduced affinity. Thus, three other mutations must also play an important role in the transition from avian to human preference. The Gal-2 and GlcNAc-3 of the avian receptor analogs in complex with the AH-1 form extensive interactions with the 220-loop, which was not observed for other HA-avian receptor complexes. For example, the backbone oxygen from G225 interacts with Gal-2, while the side chain of Q222 forms hydrogen bonds with the GlcNAc-3. The ligand is in a *cis* conformation, which is typical for complexes of avian receptors with human-infecting strains. The conformation of the avian receptor analog in complex with SH-1 closely resembles one from the complex with AH-1. For the human receptor analog, the ligand is in a *cis* conformation with L226, providing a hydrophobic base for the Gal-2 C6 atom. In complex with SH-1, only the electron density for Sia-1 was visible, a clear indication of poor binding of the 6′SLNLN. As the L266Q mutant of AH-1 could still bind the human receptors, three closely located residues (A138, V186 and P221) were able to create a hydrophobic region in the RBS, apparently alleviating the influence of Q226. For SH-1, four hydrophilic residues (S138, G186, T221 and Q226) are expected to create an unfavorable environment for the nonpolar portion of the human receptor analog, preventing binding to this HA. For AH-1, the distances between the 130-loop and the 220-loop of the RBS are larger by around 1.5 Å when compared with SH-1, which also indicates an adaptation for human receptors.

Despite the fact that the A/Shanghai/2/2013 (SH-2) strain also bears the Q226L mutation with hydrophobic amino acids at positions 138, 186 and 221, it exhibits only limited binding to human receptors, which is abolished after mutating L226Q [73]. Interestingly, the crystal structure of the human receptor analog in the complex with SH-2 bears some resemblance to the conformation of this ligand with the pandemic H2. According to the authors of [75], insertion in the 150-loop with respect to H2 might be responsible for only limited binding of the human receptors. The longer 150-loop reduced the space available for the human receptors and influenced the interaction near the Gal-4 and Glc-5 with the nearby 190-helix. Thus, similar to deletion of the 133a amino acid in the H5 subtype, deletion in the 150-loop might be necessary for more effective binding of the human receptors by H7. As H7 is also less thermostable, a mutation of similar effect to T318I, which was observed in H5, would be required for effective transmissibility between humans. Concluding, additional mutations are still needed for the H7 subtype virus to achieve a heightened specificity for human receptors, and the current pandemic potential of the H7 virus remains questionable [70].

Several studies showed that most of the recently isolated H9N2 IAVs in China have acquired the mammalian type amino acid at position 226 [76, 77]. To date, two mutations, Q226L and I155T in the HA have been linked to the determination of receptor specificity of this subtype, with a lesser role associated with the amino acids at positions 183, 190 and 228 [77]. The strain bearing the mutations T189A in HA1 and G192R in HA2, together with the already present L226, E190, A189 and H183 residues, was able to transfer via aerosol droplet transmission in ferrets. Additional mutations in NA, PB2 and M1 proteins were also identified for this strain; however, only mutations of the surface proteins were essential. This strain, however, has not been identified in nature [78]. Additional analysis showed that viruses carrying V190 are more efficiently replicated in mice, despite the fact that Val at position 190 in the HA did not affect the H9N2 virus receptor binding specificity [79]. How these mutations affect the binding of human- and avian-type analogs is still unknown, as no structures for these mutants have been proposed.

Despite the evidence of zoonotic infections with H10N8 strains, it was recently shown that human-isolated H10N8 HA retains a strong preference for avian receptor analogs [80]. Introduction of a typical set of mutations that enabled adaption of receptor specificity for the human pandemic viral subtype

failed to switch receptor preference. For example, the Q226L and G228S combination only resulted in an inability of the H10 to recognize both the human and avian receptor analogs. In recent work [7], an H10 mutant with preference shifted to human-type receptors was identified. The importance of the 150-loop was highlighted, where its length is one of the main features distinguishing the H10, H7 and H15 phylogenetic group from other HA subtypes. The loop is two amino acids longer (K158a and G158b in case of A/Jiangxi/IPB13a/2013) with respect to H3 and is located near the exit point for the human receptor observed in other HA. This insertion might hamper the binding of human type receptor to the H7, H10 and H15 subtypes, as was already suggested for the H7N9 A/Shanghai/2/2013 strain. Although substitution of K158a to a smaller residue (A or G) is insufficient to allow binding of the human receptors, further modifications (Q226L and G228S) resulted in significantly increased avidity to the human-type receptors with a simultaneously reduced binding to the avian-type receptors. Introduction of an addition D193T mutation resulted in a further increase in binding to human-type receptors. This mutant was also confirmed to bind to human trachea tissue.

The crystal structure of the mutated H10 and wild type (wt) H10 (A/Jiangxi/IPB13a/2013) with human receptor analogs (6′SLNLN and 6′SLN, respectively) reveals that in both cases, the ligands are in a *cis* conformation, although the Q226L and G228S mutations resulted in the Sia-1 being placed around 1 Å closer to the 130-loop. The second difference is the orientation of the GlcNAc-3 sugar which, in the mutated protein, points toward the 190-helix; a behavior observed for complexes of 6′SLNLN with human pandemic strains. For wt H10, the human ligand leaves the RBS closer to the 130-loop. Thus, the mutations in the 220-loop create more favorable binding interactions with the hydrophobic face of the human receptor and simultaneous deletion in the 150-loop allow the human receptor analog to leave RBS next to the 190-helix. In contrast to the *trans* conformation in the complex with the wt H10, the 3′SLN binds to the mutated H10 in a *cis* conformation. In complex with the mutated H10 the avian receptor analog leaves RBS closer to 220-loop when compared with wt H10, and its Sia-1 does not make hydrophobic interactions with amino acids from the 220-loop because of Q226L and G228S, resulting in weaker binding of 3′SLN. The extension in the 150-loop characteristic of H10 HAs seems to have no effect on avian receptors binding.

In 2017, a structure of the H15 subtype HA was proposed for the first time [81]. Presently, no zoonotic infections with viruses carrying this HA subtype have been reported, most likely because the virus is circulating primarily among shorebirds in Australia (a rather limited distribution). The crystal structure of the HA from the A/shearwater/WA/2576/1979 strain closely resembles that of the H7 subtype, as expected from the phylogenetic analysis of HAs. The RMSD between the H7 and H15 monomers is around 1.1 Å. A superposition of the H15 and H7 HA-trimers indicates that the head domain of the H15 is wider, adopting a more open conformation, which might promote binding of human receptors. The H15N7 virus retains a strong preference for avian receptors with the typical avian-infecting amino acids at key positions: E190, G225, Q226 and G228. Similar to H7 and H10 subtypes, an insertion in the 150-loop is present, which might hamper the binding of the human-type receptors. The unique feature of this subtype is the large insertion in the 260-loop, which is, however, outside the RBS; thus, it is unlikely that it would affect the binding of sialylated glycans. The authors obtained the structure of H15 with 3′SLN, an analog of the avian receptors. The 3′SLN is in a *trans* conformation and creates typical contacts that are observed in other subtypes with the key role of Q226. Unfortunately, the structure of the protein with the human receptor analog was not reported. The adaptive mutations characteristic of other subtypes did not allow a switch of preference to the human-type receptors; E190D and G225D resulted in no activity toward avian receptors, while the Q226L and G228S mutants retain the ability to bind avian receptors, and gained weak binding towards the long, branched, human receptor analogs. How these mutations affect the conformation of 3′SLN has not been determined.

Summary

In this work, provided a brief summary of our current knowledge on the adaptive mutations that allow IAV to switch its HA preference from avian receptors to human. IAV is unique among other influenza viruses (type B or C) as it circulates primarily outside human populations. Thus, specific mutations in a given HA strain's RBS are believed to be one of the perquisites for a strain to gain the ability to transmit between humans. The RBSs for all HA are strikingly similar with all key structural elements present in all HA subtypes. These elements are the helix 190, loop-130, loop-150 and loop-220. Remarkably, even the length of these elements is nearly identical between subtypes. Some notable exceptions are two additional residues in the 150-loop for the H7, H10 and H15 subtypes, or a variable length of 130-loop.These relatively small changes often have a profound impact on the receptor specificity of a given subtype. For example, to switch the preference of the HA from A/Jiangxi/IPB13a/2013 (H10N8) strain, K158a (located at the 150-loop) had to be mutated to alanine or glycine.

The amino acid composition of the RBS varies significantly between subtypes with only four residues Y98, W153, H183 and Y195 conserved between all the subtypes. These residues form the 'floor' of the RBS and interact directly with the Sia-1 of the bound glycans. The other residues are not universally conserved, but rather characteristic of the specific strain. For example, most avian-infecting strains carry Q226, which provides a favorable environment for ligands with the α2,3 bond and Sia-1 and Gal-2 connected in *trans*. Then, both oxygens from the carboxyl group (on Sia-1) and the oxygen from the glyosidic bond can interact with the hydrophilic side chain of Q226. On the other hand L226 introduces a hydrophobic side chain, which disfavors these interactions and promotes ligands with the α2,6 bond, which can interact with residue L226 via the C6 atom of Gal-2. Q226 is sometimes observed in strains with the human receptor preferring HA; however, the hydrophobic effect of the leucine residue can be emulated by residues at 138, 186 and 221 as in the case of some H7 strains. This is only one of many examples of correlated mutations required for HA to achieve different receptor preference.

The HA protein can also be found in other influenza viruses (like type B or C). Contrary to the situation observed for IAV, influenza B virus (IBV) strains are not assigned into subtypes based on the properties of the HA and NA proteins but rather into one of the two major linages circulating in the human population—B/Victoria or B/Yamagata [82]. IBV can cause seasonal epidemics, but unlike IAV, humans are its main reservoir. Thus, no adaptive mutations are expected to occur, as no interspecies barrier needs to be crossed. Possibly because of that fact, the typical mutations in HA from IBV of both linages are usually observed outside the RBS, near the potential antigenic sites [83]. When compared with IAV, the naturally occurring IBV showed a lower binding affinity for synthetic glycans and natural receptors on the cell surface. The IBV binding affinity for human receptors can be restored to

IAV-levels with a single mutation F95Y (B/Hong Kong/8/73 numbering); however, its implication to the pathogenicity and evolution of IBV are still unresolved [84].

For the influenza C virus (ICV), there is no 'typical' HA protein. Instead, one external protein, called HEF, combines in a single protein both the roles of HA and NA. Furthermore, HEF recognizes compounds terminated with N-acetyl-9-O-acetylneuraminic acid instead of N-acetylneuraminic acid like HA of IAV and IBV. The connection type between the terminal and penultimate saccharides is also not a factor during ICV infection. ICV circulates predominantly among humans, a characteristic it shares with IBV; however, contrary to IBV, it causes only mild symptoms. Recently, a HEF protein from ICV was the subject of a comprehensive review [85].

Key Points

- HA is a transmembrane protein, which allows the IAV to attach to the host cell by recognizing specific glycans.
- The main difference distinguishing IAV circulating in human and avian species is the preferential recognition of glycan terminated with the α2,6 linkage by the former and α2,3 linkage by the latter.
- Modifications in the HA RBS can lead to switch in preference to specific glycan, and this may lead to a pandemic if the HA of avian-infecting virus switch its preference to human receptors.
- E190D and G225D mutations for HA from Subtype 1 and Q226L and G228S mutations for HAs from Subtypes 2 and 3 changed the preference from human- to avian-type receptors.
- Several different sets of mutations were described for other subtypes, which highlights the threat to human populations posed by IAV currently circulating among avian reservoirs.

Funding

This work was supported by grants from the Polish National Science Centre (grant numbers 2014/15/B/ST6/05082, 2013/09/B/NZ2/00121); Foundation for Polish Science (TEAM to DP); 1U54DK107967-01 'Nucleome Positioning System for Spatiotemporal Genome Organization and Regulation' within 4DNucleome NIH program.

References

1. Webster RG, Bean WJ, Gorman OT. Evolution and ecology of influenza A viruses. *Microbiol Rev* 1992;**56**(1):152–79.
2. Wu Y, Tefsen B, Shi Y, *et al*. Bat-derived influenza-like viruses H17N10 and H18N11. *Trends Microbiol* 2014;**22**(4):183–91.
3. Yamayoshi S, Watanabe M, Goto H, *et al*. Identification of a novel viral protein expressed from the PB2 segment of influenza A virus. *J Virol* 2015;**90**(1):444–56.
4. Jagger BW, Wise HM, Kash JC, *et al*. An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science* 2012;**337**(6091):199–204.
5. Król E, Rychłowska M, Szewczyk B. Antivirals–current trends in fighting influenza. *Acta Biochimica Polonica* 2014;**61**(3):495–504.
6. de Graaf M, Fouchier RA. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. *EMBO J* 2014;**33**(8):823–41.
7. Tzarum N, de Vries RP, Peng W, *et al*. The 150-loop restricts the host specificity of human H10N8 influenza virus. *Cell Rep* 2017;**19**(2):235–45.
8. Matrosovich M, Stech J, Klenk HD. Influenza receptors, polymerase and host range. *Rev Sci Tech* 2009;**28**(1):203–17.
9. Zambon M. Lessons from the 1918 influenza. *Nat Biotechnol* 2007;**25**(4):433–4.
10. Miotto O, Heiny AT, Albrecht R, *et al*. Complete-proteome mapping of human influenza A adaptive mutations: implications for human transmissibility of zoonotic strains. *PLoS One* 2010;**5**(2):e9025.
11. Lin YP, Xiong X, Wharton SA, *et al*. Evolution of the receptor binding properties of the influenza A(H3N2) hemagglutinin. *Proc Natl Acad Sci USA* 2012;**109**(52):21474–9.
12. Neumann G, Noda T, Kawaoka Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* 2009;**459**(7249):931–9.
13. Xu R, McBride R, Nycholat CM, *et al*. Structural characterization of the hemagglutinin receptor specificity from the 2009 H1N1 influenza pandemic. *J Virol* 2012;**86**(2):982–90.
14. Girard MP, Tam JS, Assossou OM, *et al*. The 2009 A (H1N1) influenza virus pandemic: a review. *Vaccine* 2010;**28**(31):4895–902.
15. Smith GJ, Vijaykrishna D, Bahl J, *et al*. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 2009;**459**(7250):1122–5.
16. Wei SH, Yang JR, Wu HS, *et al*. Human infection with avian influenza A H6N1 virus: an epidemiological analysis. *Lancet Respir Med* 2013;**1**(10):771–8.
17. Yang H, Carney PJ, Donis RO, *et al*. Structure and receptor complexes of the hemagglutinin from a highly pathogenic H7N7 influenza virus. *J Virol* 2012;**86**(16):8645–52.
18. Puzelli S, Rossini G, Facchini M, *et al*. Human infection with highly pathogenic A(H7N7) avian influenza virus, Italy, 2013. *Emerg Infect Dis* 2014;**20**(10):1745–9.
19. Huang Y, Li X, Zhang H, *et al*. Human infection with an avian influenza A (H9N2) virus in the middle region of China. *J Med Virol* 2015;**87**(10):1641–8.
20. Kilbourne ED. Influenza pandemics of the 20th century. *Emerg Infect Dis* 2006;**12**(1):9–14.
21. Das K, Aramini JM, Ma LC, *et al*. Structures of influenza A proteins and insights into antiviral drug targets. *Nat Struct Mol Biol* 2010;**17**(5):530–8.
22. Kadam RU, Wilson IA. Structural basis of influenza virus fusion inhibition by the antiviral drug Arbidol. *Proc Natl Acad Sci USA* 2017;**114**(2):206–14.
23. Russell RJ, Kerry PS, Stevens DJ, *et al*. Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. *Proc Natl Acad Sci USA* 2008;**105**(46):17736–41.
24. Wu X, Sun Q, Zhang C, *et al*. Progress of small molecular inhibitors in the development of anti-influenza virus agents. *Theranostics* 2017;**7**(4):826–45.
25. Eng CL, Tong JC, Tan TW. Predicting host tropism of influenza A virus proteins using random forest. *BMC Med Genomics* 2014;**7**(Suppl 3):1755–8794.
26. Eng CLP, Tong JC, Tan TW, Chan MCW. Distinct host tropism protein signatures to identify possible zoonotic influenza A viruses. *PLoS One* 2016;**11**(2):e0150173.
27. Harrison SC. Viral membrane fusion. *Nat Struct Mol Biol* 2008;**15**(7):690–8.
28. Carr CM, Chaudhry C, Kim PS. Influenza hemagglutinin is spring-loaded by a metastable native conformation. *Proc Natl Acad Sci USA* 1997;**94**(26):14306–13.
29. White JM, Whittaker GR. Fusion of enveloped viruses in endosomes. *Traffic* 2016;**17**(6):593–614.

30. Smrt ST, Lorieau JL. Membrane fusion and infection of the influenza hemagglutinin. *Adv Exp Med Biol* 2017;**966**:37–54.
31. Worch R. The helical hairpin structure of the influenza fusion peptide can be seen on a hydrophobic moment map. *FEBS Lett* 2013;**587**(18):2980–3.
32. Worch R, Krupa J, Filipek A, *et al*. Three conserved C-terminal residues of influenza fusion peptide alter its behavior at the membrane interface. *Biochim Biophys Acta* 2017;**1861**(2): 97–105.
33. Gamblin SJ, Skehel JJ. Influenza hemagglutinin and neuraminidase membrane glycoproteins. *J Biol Chem* 2010;**285**(37): 28403–9.
34. Ma W, Garcia-Sastre A, Schwemmle M. Expected and unexpected features of the newly discovered bat influenza A-like viruses. *PLoS Pathogens* 2015;**11**(6):e1004819.
35. Xiong X, McCauley JW, Steinhauer DA. Receptor binding properties of the influenza virus hemagglutinin as a determinant of host range. *Curr Top Microbiol Immunol* 2014;**385**:63–91.
36. Wilson IA, Skehel JJ, Wiley DC. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. *Nature* 1981;**289**(5796):366–73.
37. Zhang Y, Aevermann BD, Anderson TK, *et al*. Influenza research database: an integrated bioinformatics resource for influenza virus research. *Nucleic Acids Res* 2017;**45**:D466–74.
38. Mazzocco G, Lazniewski M, Migdał P, *et al*. 3DFlu: database of sequence and structural variability of the influenza hemagglutinin at population scale. *Database* 2016;**2016**:baw130.
39. Ambroggio XI, Dommer J, Gopalan V, *et al*. HASP server: a database and structural visualization platform for comparative models of influenza A hemagglutinin proteins. *BMC Bioinformatics* 2013;**14**(1):197.
40. Nicholls JM, Chan RW, Russell RJ, *et al*. Evolving complexities of influenza virus and its receptors. *Trends Microbiol* 2008; **16**(4):149–57.
41. Walther T, Karamanska R, Chan RW, *et al*. Glycomic analysis of human respiratory tract tissues and correlation with influenza virus infection. *PLoS Pathogens* 2013;**9**(3):e1003223.
42. Franca M, Stallknecht DE, Howerth EW. Expression and distribution of sialic acid influenza virus receptors in wild birds. *Avian Pathol* 2013;**42**:60–71.
43. Xiong X, Xiao H, Martin SR, *et al*. Enhanced human receptor binding by H5 haemagglutinins. *Virology* 2014;**456–457**: 179–87.
44. Stray SJ, Cummings RD, Air GM. Influenza virus infection of desialylated cells. *Glycobiology* 2000;**10**(7):649–58.
45. Arcanjo AC, Mazzocco G, de Oliveira SF, *et al*. Role of the host genetic variability in the influenza A virus susceptibility. *Acta Biochimica Polonica* 2014;**61**(3):403–19.
46. Zhao N, Martin BE, Yang CK, *et al*. Association analyses of large-scale glycan microarray data reveal novel host-specific substructures in influenza A virus binding glycans. *Sci Rep* 2015;**5**(1):15778.
47. Murphy BR, Wood FT, Massicot JG, *et al*. Temperature-sensitive mutants of influenza virus. XV. The genetic and biological characterization of a recombinant influenza virus containing two ts lesions produced by mating two complementing, single lesion ts mutants. *Virology* 1978;**88**:231–43.
48. Baum LG, Paulson JC. The N2 neuraminidase of human influenza virus has acquired a substrate specificity complementary to the hemagglutinin receptor specificity. *Virology* 1991; **180**(1):10–15.
49. Ginting TE, Shinya K, Kyan Y, *et al*. Amino acid changes in hemagglutinin contribute to the replication of oseltamivir-resistant H1N1 influenza viruses. *J Virol* 2012;**86**(1):121–7.
50. Byrd-Leotis L, Cummings RD, Steinhauer DA. The interplay between the host receptor and influenza virus hemagglutinin and neuraminidase. *Int J Mol Sci* 2017;**18**.
51. Ni F, Kondrashkina E, Wang Q. Structural and functional studies of influenza virus A/H6 hemagglutinin. *PLoS One* 2015; **10**(7):e0134576.
52. Kash JC, Taubenberger JK. The role of viral, host, and secondary bacterial factors in influenza pathogenesis. *Am J Pathol* 2015;**185**(6):1528–36.
53. Gamblin SJ, Haire LF, Russell RJ, *et al*. The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* 2004;**303**(5665):1838–42.
54. Yang H, Carney P, Stevens J. Structure and receptor binding properties of a pandemic H1N1 virus hemagglutinin. *PLoS Curr* 2010;**2**:RRN1152.
55. Pappas C, Yang H, Carney PJ, *et al*. Assessment of transmission, pathogenesis and adaptation of H2 subtype influenza viruses in ferrets. *Virology* 2015;**477**:61–71.
56. Liu J, Stevens DJ, Haire LF, *et al*. Structures of receptor complexes formed by hemagglutinins from the Asian Influenza pandemic of 1957. *Proc Natl Acad Sci USA* 2009;**106**(40): 17175–80.
57. Ha Y, Stevens DJ, Skehel JJ, *et al*. X-ray structure of the hemagglutinin of a potential H3 avian progenitor of the 1968 Hong Kong pandemic influenza virus. *Virology* 2003;**309**(2):209–18.
58. Weis WI, Brunger AT, Skehel JJ, *et al*. Refinement of the influenza virus hemagglutinin by simulated annealing. *J Mol Biol* 1990;**212**:737–61.
59. Zhu X, Viswanathan K, Raman R, *et al*. Structural basis for a switch in receptor binding specificity of two H5N1 hemagglutinin mutants. *Cell Rep* 2015;**13**(8):1683–91.
60. Stevens J, Blixt O, Paulson JC, *et al*. Glycan microarray technologies: tools to survey host specificity of influenza viruses. *Nat Rev Microbiol* 2006;**4**(11):857–64.
61. de Vries RP, Zhu X, McBride R, *et al*. Hemagglutinin receptor specificity and structural analyses of respiratory droplet-transmissible H5N1 viruses. *J Virol* 2014;**88**(1):768–73.
62. Xiong X, Coombs PJ, Martin SR, *et al*. Receptor binding by a ferret-transmissible H5 avian influenza virus. *Nature* 2013; **497**(7449):392–6.
63. Imai M, Watanabe T, Hatta M, *et al*. Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* 2012;**486**(7403):420–8.
64. Herfst S, Schrauwen EJ, Linster M, *et al*. Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* 2012; **336**(6088):1534–41.
65. Chen LM, Blixt O, Stevens J, *et al*. *In vitro* evolution of H5N1 avian influenza virus toward human-type receptor specificity. *Virology* 2012;**422**(1):105–13.
66. Zhang W, Shi Y, Lu X, *et al*. An airborne transmissible avian influenza H5 hemagglutinin seen at the atomic level. *Science* 2013;**340**(6139):1463–7.
67. Tharakaraman K, Raman R, Viswanathan K, *et al*. Structural determinants for naturally evolving H5N1 hemagglutinin to switch its receptor specificity. *Cell* 2013;**153**(7):1475–85.
68. Tzarum N, de Vries RP, Zhu X, *et al*. Structure and receptor binding of the hemagglutinin from a human H6N1 influenza virus. *Cell Host Microbe* 2015;**17**(3):369–76.
69. Wang F, Qi J, Bi Y, *et al*. Adaptation of avian influenza A (H6N1) virus from avian to human receptor-binding preference. *EMBO J* 2015;**34**(12):1661–73.
70. Nailwal H, Kamra K, Lal SK. H7N9: a killer in the making or a false alarm? *Can J Microbiol* 2014;**60**(7):425–9.

71. Belser JA, Creager HM, Sun X, *et al*. Mammalian pathogenesis and transmission of H7N9 influenza viruses from three waves, 2013-2015. *J Virol* 2016;**90**(9):4647–57.
72. Zhang Q, Shi J, Deng G, *et al*. H7N9 influenza viruses are transmissible in ferrets by respiratory droplet. *Science* 2013; **341**(6144):410–14.
73. Xu R, de Vries RP, Zhu X, *et al*. Preferential recognition of avian-like receptors in human influenza A H7N9 viruses. *Science* 2013;**342**(6163):1230–5.
74. Shi Y, Zhang W, Wang F, *et al*. Structures and receptor binding of hemagglutinins from human-infecting H7N9 influenza viruses. *Science* 2013;**342**(6155):243–7.
75. Xiong X, Martin SR, Haire LF, *et al*. Receptor binding by an H7N9 influenza virus from humans. *Nature* 2013;**499**(7459): 496–9.
76. Wan H, Perez DR. Amino acid 226 in the hemagglutinin of H9N2 influenza viruses determines cell tropism and replication in human airway epithelial cells. *J Virol* 2007;**81**(10): 5181–91.
77. Li X, Shi J, Guo J, *et al*. Genetics, receptor binding property, and transmissibility in mammals of naturally isolated H9N2 avian influenza viruses. *PLoS Pathogens* 2014;**10**(11): e1004508.
78. Sorrell EM, Wan H, Araya Y, *et al*. Minimal molecular constraints for respiratory droplet transmission of an avian-human H9N2 influenza A virus. *Proc Natl Acad Sci USA* 2009; **106**(18):7565–70.
79. Teng Q, Xu D, Shen W, *et al*. A single mutation at position 190 in hemagglutinin enhances binding affinity for human type sialic acid receptor and replication of H9N2 avian influenza virus in mice. *J Virol* 2016;**90**(21):9806–25.
80. Zhang H, de Vries RP, Tzarum N, *et al*. A human-infecting H10N8 influenza virus retains a strong preference for avian-type receptors. *Cell Host Microbe* 2015;**17**:377–84.
81. Tzarum N, McBride R, Nycholat CM, *et al*. Unique structural features of influenza H15 HA. *J Virol* 2017;**91**:e00046–17.
82. Ni F, Kondrashkina E, Wang Q. Structural basis for the divergent evolution of influenza B virus hemagglutinin. *Virology* 2013;**446**(1–2):112–22.
83. Vijaykrishna D, Holmes EC, Joseph U, *et al*. The contrasting phylodynamics of human influenza B viruses. *Elife* 2015;**4**:e05055.
84. Ni F, Mbawuike IN, Kondrashkina E, *et al*. The roles of hemagglutinin Phe-95 in receptor binding and pathogenicity of influenza B virus. *Virology* 2014;**450–451**:71–83.
85. Wang M, Veit M. Hemagglutinin-esterase-fusion (HEF) protein of influenza C virus. *Protein Cell* 2016;**7**(1):28–45.