Glycosylation of the influenza A virus Hemagglutinin protein

Melissa Boeijen, 3219674

Abstract

Influenza is a dangerous virus which has been the cause of many epidemics in the past and may create many more health issues in the future. Each year, high risk groups of the population are vaccinated against the influenza virus, however the antibodies generated against the vaccine might not fit the viral strain that is circling at that moment. This can be due to the form of hemagglutinin. The glycoprotein hemagglutinin is located on the envelope of influenza virus particles and can have different glycosylation forms. For antibody recognition the length of the glycans and the location of the glycosylation site are important. In this review an overview is provided of the glycosylation of the hemagglutinin protein and how it affects the functional properties of hemagglutinin. Lastly, it will be discussed how this knowledge may contribute to the development of influenza virus vaccines.

Introduction

Influenza is an enveloped RNA virus and belongs to the *Orthomyxoviridae* family [1, 2]. During the 20th century the influenza A virus caused several pandemics, one of them was the 'Spanish' flu (1918-1919), which killed approximately 20 million to 40 million people [3-6]. In the beginning of the 21st century another pandemic came to pass: the Mexican flu [7], and this year an avian influenza known as H7N9, infected humans in China [8].

The virus particles of influenza A consists of three major components namely: the core, the matrix and the envelope, see **figure 1**. The core contains eight negative RNA strands, which are situated in ribonucleocapsids (vRNPs), and nucleoprotein. Furthermore, the nuclear export protein is also present with three polymerase proteins [1, 9, 10].

The matrix protein (M1) is located between the core and the envelope of the virus particle [1]. M1 determines the shape of the virus particle and plays multiple roles in the viral replication cycle [11].

The envelope is comprised of a lipid bilayer containing three proteins: M2, neuraminidase (NA) and hemagglutinin (HA) [1, 4, 12-15]. M2 is a type III transmembrane protein and creates an ion channel to uncoat the viral particles. While NA is a type II transmembrane protein. Approximately 17% of the envelope consist of this glycoprotein and forms homogenous tetramers. It has the function to cleave the binding of the sialic acid residues that are on the cell surface of the host cell and thereby releasing the newly assembled virus particles [1, 9, 13, 16]. Additionally, with 80% HA is the most abundant envelope glycoprotein and contains approximately 566 amino acids, depending on the strain [9, 17]. Furthermore, HA is a type I transmembrane protein and it forms homogenous trimers, see **figure 2** [17, 18]. A HA monomer (HA0) consists of two parts, HA1 (± 36 kDa) and HA2 (± 27 kDa), which are linked by a disulfide bond [17, 19, 20]. In addition, viral entry is mediated by HA, as it binds sialic acid residues on the receptors of the host cell [14, 16, 21-23]. Ultimately, HA mediates membrane fusion [22-24].

During maturation in the cell the protein HA is glycosylated. The aim of this review is to acquire insight into how glycosylation affects the functional properties of HA. Furthermore, it will be discussed how this knowledge may contribute in the development of influenza virus vaccines.

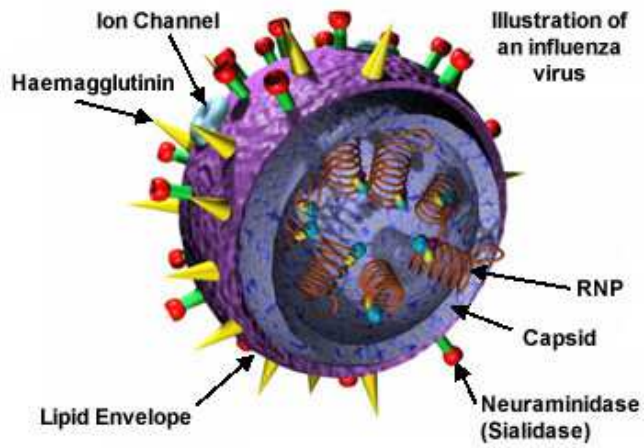

Figure 1. Schematic presentation of an influenza virus particle. Retrieved from http://www.wvdhhr.org/labservices/labs/virology/influenza.cfm on 2013/05/30

Figure 2. HA in trimeric conformation. Each color shows a HA monomer. Retrieved from http://biology.kenyon.edu/BMB/Chime2/2005/Cerchiara-Holsberry/FRAMES/start.htm on 2013/07/06

Glycosylation of hemagglutinin

Hemagglutinin structure and domains

Since HA and NA are the most abundant proteins on the envelope of the virus particles of influenza A, the viral strains are classified based on the antibody response on these proteins. For example, in H1N1, H stands for HA and N stands for NA [2]. There are seventeen different HAs and nine NAs [15, 18, 25, 26] from which humans can be infected with H1, H2 and H3. Additionally, other common human infections are N1 and N2 [2, 13, 22, 27]. H3N2 and H1N1 are nowadays the viral strains that are most present, although most seasonal epidemics are caused by H3N2. Since H3N2 strains are more often present than H1N1 strains, they adapt more often to avoid immunity of the host. The diversity in H3N2 strains is therefore higher than in the H1N1 strains [28-31].

The different HAs can be divided into two groups: group 1 (which is also called the H1 group) and group 2 (which is also named the H3 group), see **figure 3** [28, 32, 33]. These groups are categorized 'based on their antigenic properties and their major structural features' [28].

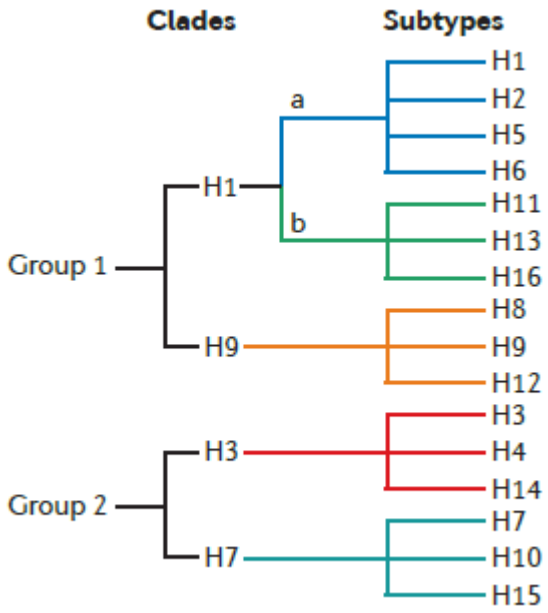

Figure 3. Influenza virus strains are divided into two groups [28].

As already discussed in the introduction, HA mediates viral entry by receptor binding and membrane fusion between the virus particle and the endosomal membrane [14, 16, 21, 22, 24]. HA consists of two structural parts: the stem region, which consists of HA2, and the head region, which consists of HA1 [22, 27, 34-36]. As can be seen in **figure 4**, the stem domain of HA consist mainly of α-helixes, while the head domain is primarily build of β-strands. Receptor binding is accomplished by a shallow cavity in the head region of HA [22, 27, 30, 35] and is well conserved in all subtypes [32]. While the stem region mediates membrane fusion [24].

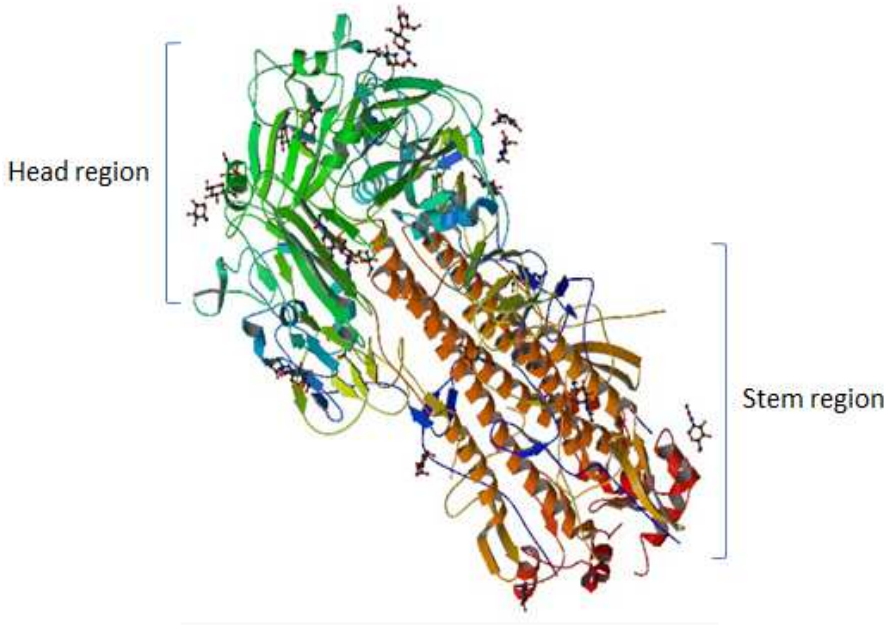

Figure 4. HA0 is divided into two regions: the head and stem regions. Retrieved from http://www.rcsb.org/pdb/explore/explore.do?structureId=2VIU on 2013/06/04

The reason why some influenza viruses only infect humans and others solely infect birds is caused by the binding of the receptor with HA. These receptors contain sialic acids, which is present in high concentrations on glycolipids and glycoproteins [1]. Every type of influenza strain has its own preference of binding to the sialic acid: α(2, 6) – linkage and α(2, 3) – linkage. For the α(2,3) – linkage it means that the sialic acid is linked with the second carbon atom of the hexane ring to the sugar group. This sugar group is connected to the sialic acid with the third carbon atom in the hexose, see **figure 5**. The same applies for the α(2,6) – linkage: the sialic acid is linked with the second carbon atom of the hexane ring to the sugar group and the sugar group is connected to the sialic acid with the sixth carbon atom in the hexose [22, 37]. The difference in binding specificity is due to the different species that can be infected with influenza: human influenza has a preference for α(2, 6) – linkage and avian influenza prefers α(2, 3) – linkage, while swine influenza can be infected with both types of linkages [22]. The reason for the fact that human receptors bind with the α(2, 6) – linkage to sialic acids is that human lungs and other parts of the respiratory tract are mostly covered with this type of receptor [22, 38].

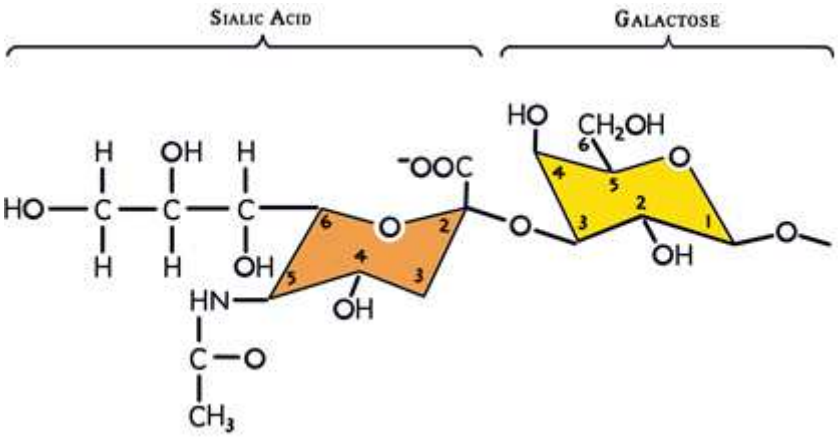

Figure 5. α(2,3) – linked sialic acid to galactose. Retrieved from http://www.virology.ws/2009/05/05/influenza-virus-attachment-to-cells-role-of-different-sialic-acids/ on 2013/06/11

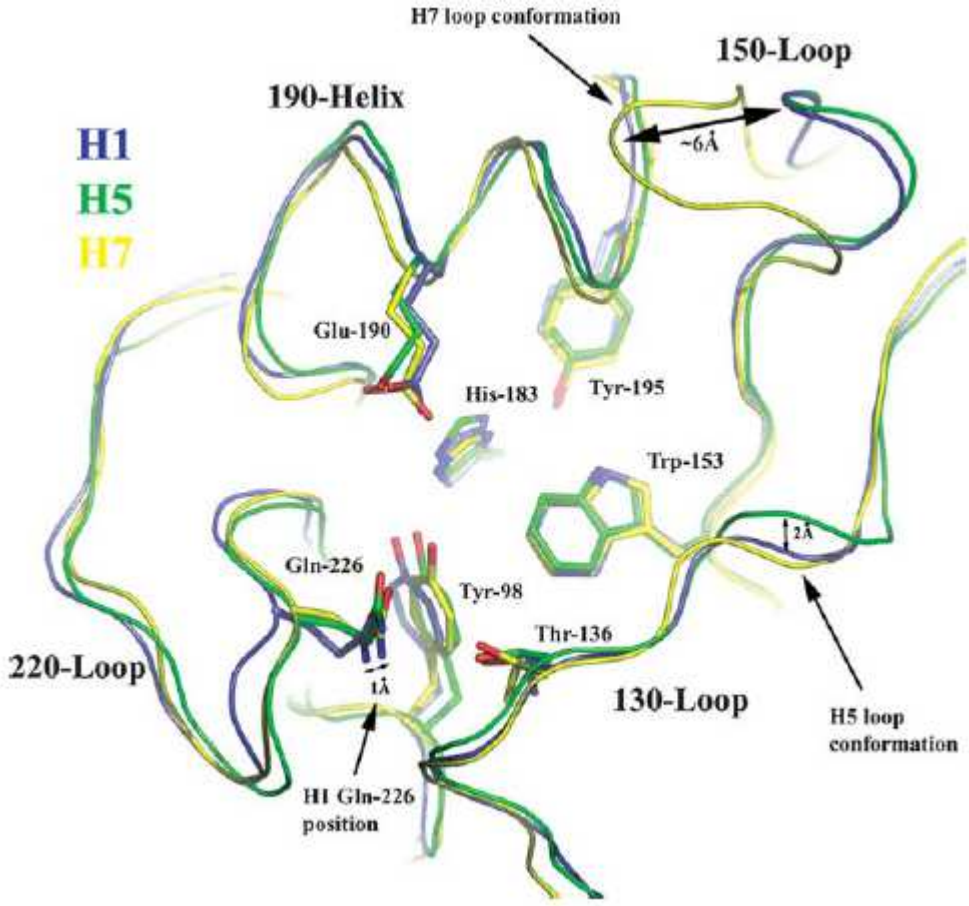

Figure 6. Overlay of the region where receptor binding takes place of different HA subtypes. The loop-helix-loop motive is present. Only the (avian) H7 subtype has an extra loop (150-loop). Additionally, the 130-loop of the (avian) H5 subtype is shifted [27].

For both linkages receptor binding is mediated by a loop-helix-loop motive. The first loop is located at amino acids 135-138 and is called the 130-loop. The helix can be found at amino acids 190-198 and is named the 190-helix. In addition, the last loop is called the 220-loop and is located at amino acids 221-228, see **figure 6** [27, 39]. Furthermore, researchers discovered that certain amino acids are well conserved, like the tyrosine at position 98, tryptophan located at residue 153, histidine at position 183, leucine located at residue 194 and the tyrosine at position 195. They form hydrogen bonds with the sialic acids [22, 35, 39, 40] and are thus responsible for the three dimensional structure of HA. This way they control HA binding to sialic acids as well as the transfer of influenza between the species [27, 35]. Furthermore, certain amino acids play a role in determining which linkage is preferred and even small changes as two mutated amino acids can make the difference in the preference of the linkage[22, 23, 27, 37, 41]. In H1, HA receptor binding is controlled by amino acids 190, 226 and 228. Especially residue 190 is important for binding to an avian host. In contrast to HA of H3; in this case residues 226 and 228 play a key role in the α(2, 3) – linkage [41]. Which are leucine and serine in human influenza for residues 226 and 228, while avian influenza strains have a glutamine and glycine at these positions. The different amino acids in humans create a distance between the receptor and HA. Besides, the amino acid at position 226 does not bind sialic acids, but it is important for the structural conformation of the binding site of HA [22].

Hu [23] investigated the H1N1 strain that caused the pandemic of 2009. This strain binds with the α(2, 6) – linkage as well as the α(2, 3) – linkage, which is not common for the annual influenza plague. Hu[23] aligned the H1N1 strain of 2009 with swine H1N1 and avian H5N1 and it was discovered that there are conserved residues; in the strain of H1N1 of 2009 these were amino acids 106-130, 150-174 and 191-221, see **figure 7**. In addition, mutations in amino acids 182 and 192 in HA of H5 switch the binding affinity from an α(2, 3) – linkage to an α(2, 6) – linkage. The researcher found that the H1N1 strain of 2009 had the aspartic acid at position 190 and the aspartic acid at position 225, which imparts a human binding. In avian influenza there is glutamic acid at position 190 and the mutation of glutamic acid to aspartic acid results in the loss of binding for an α(2, 3) – linkage, however this is undone by the lysine at position 145 in the H1N1 strain of 2009. Further, the H1N1 strain of 2009 appeared to have more than two loops: a 130-loop, 140-loop, 150-loop and 220-loop were signaled and all of these loops have a lysine residue, which support the binding of an α(2, 3) – linkage, next to the α(2, 6) – linkage which is caused by the aspartic acid [23].

Figure 7. Sequence alignment of HA1 of the H1N1 strain of 2009, swine H1N2 and avian H5N1. Red indicates binding sites. The bold amino acids are the edges of the binding sites. The conserved amino acid residues in HA1 are: 91, 150, 152, 180, 187, 191 and 192 [23].

After HA binds to the sialic acid receptor the virus particle internalizes into the host cell by endocytosis [1, 22, 42] and the environment of the virus particle gets acidified (pH is about 5.0). The virus particle is then uncoated with the help of M2 and the fusion of the membranes is activated by the cleavage of a loop in HA0. The cleavage of HA0 results in the formation of HA1 and HA2, which are linked by a disulfide bond [12, 16, 17, 22, 33, 42-45]. The cleavage of HA is the activating key for membrane fusion and therefore it is also an intracellular marker for transport [13, 30, 33, 37, 44]. Moreover, if HA0 is not cleaved, virus particles appear to be non-infectious [1, 9, 22, 33]. The split of HA0 into HA1 and HA2 (in all subtypes of HA), is located at a basic residue. However, the region where cleavage takes places is different for the HA subtypes, depending if they are High Pathogenic Avian Influenza (HPAI) or Low Pathogenic Avian Influenza (LPAI) [30]. HA0 in LPAI is cleaved 'at a conserved arginine residue by trypsin-like proteases' [13], for example by tryptase Clara. These proteases can be found in humans mainly in the respiratory tract. While HA0 in HPAI is cleaved at multi-basic amino acids. They can be cleaved by ubiquitous proteases. These proteases are present at multiple places in the body and not

only at the respiratory tract [1, 13, 33, 37, 44]. They are thus not bound to a certain area in the body, which increases the change of disease and thus making HPAI more dangerous than LPAI; it causes more deaths [1, 13, 30].

During cleavage of HA0, HA1 acquires a C-terminal end and HA2 acquires a N-terminus. This N-terminal of HA2 is also called the fusion peptide and it is the most conserved region of HA [20, 22, 30, 34, 46]. Other domains in HA2, which are responsible for membrane fusion, are 'the heptad repeat regions, transmembrane domain and the cytoplasmic tail' [20, 30, 33]. When the fusion process starts the structure of HA changes irreversibly, to give a short overview: the head region detrimerizes, the fusion peptide is no longer buried, the HA2 coiled-coil is extended and a part of the coiled-coil is transformed to a loop [47]. Prior cleavage, the fusion peptide is located in the inside of the HA trimer. However, when HA0 is cleaved it will bend outwards and anchors itself into the membrane of the endosome by hydrophobic interactions with membrane lipids. This way it connects the viral and host membranes [20, 22, 30, 34, 46]. The transmembrane domain is then located at the same side as the fusion peptide and they have a weak interaction [20, 47]. The part of the fusion peptide that is connecting to the membrane is about ten amino acids long and membrane fusion is only occurring when at least two or more of these amino acids are present. Moreover, the amino acids close to the fusion peptide affect the structure of HA and so the fusion process [22, 47]. Certain mutations near the fusion peptide result in a more stable structure of HA, in such a case the fusion pH is lower than that of the wild type. This happens especially when amino acid at position 17 in H3 (histidine) was substituted to tyrosine, see **figure 8**. The reason for this result is that the HA structurally changes into a more thermostable conformation, while membrane fusion takes place. Apparently this form is energetically favorable [33]. However, certain mutations can also lead to a less stable structure of HA. This happens to almost all mutations in the fusion peptide that are registered in the N-

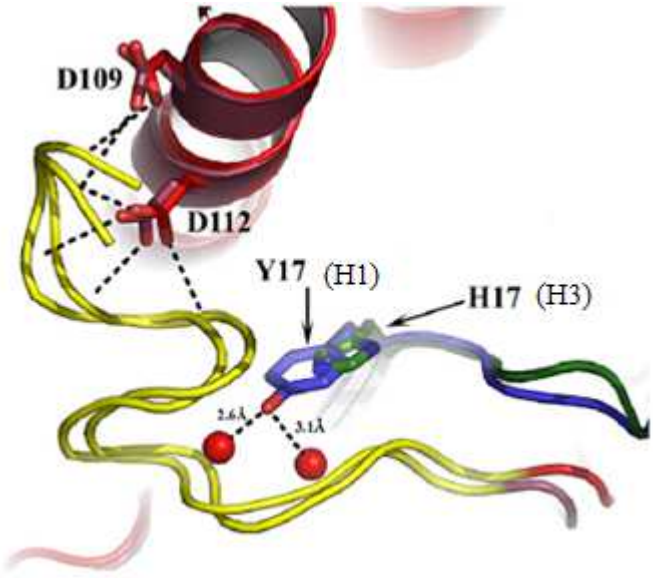

Figure 8. Structural overlay of the H1 and H3 region around the fusion peptide. Yellow is fusion peptide. The dashed lines are hydrogen bonds. The tyrosine residue in H1 forms hydrogen bonds with the fusion peptide, while the histidine of H3 does not. These hydrogen bonds are probably necessary to stabilize the HA structure at neutral pH, since histidine can change its charged states [33].

terminal part: it results in a higher fusion pH than that of the wild type [33]. With the help of the transmembrane domain of HA2 the fusion peptide is also connected to the viral membrane. This transmembrane domain also influences the depth of the fusion peptide into the membrane, which is pH dependent; at a lower pH the fusion peptide is deeper inserted into the membrane [20, 22, 30, 34]. It is thought that the transmembrane domain organizes fusion through hemifusion, since mutation in the transmembrane domain can block membrane fusion [20]. Hemifusion takes place by the formation of a stalk of the heptad repeat regions which brings the two membranes together. Then the two membranes mix their lipid compositions[20, 30], the stalk is broken by the formation of a pore to release the viral RNA into the cytosol of the host cell [30].

Although cleavage is depending on the proteases that are available, HA2 of both HA groups has conserved residues at position 51 (lysine), 109 (aspartic acid) and 112 (aspartic acid). Substitutions of one of these aspartic acids to alanine leads to an increased fusion pH. Which might be due to the point that aspartic acid at position 109 is linked to glutamine in position 105 via a water molecule. This aspartic acid is also linked to the fusion peptide by a hydrogen bond. In addition, aspartic acid at position 112 is as well linked to the fusion peptide by multiple hydrogen bonds, which can be seen in **figure 8.** Most probably these hydrogen bonds are stabilizing the HA conformation at a neutral pH. Although some residues are group specific, in HA2 amino acid 106 is in group 2 almost always histidine and in group 1 arginine or lysine. The histidine of group 2 can form hydrogen bonds with the lysine at position 51 and if the amino acid at this position is not histidine the pH at which fusion occurs is decreased. Furthermore, in group 1 the arginine at position 106 of HA2 can connect to the aspartic acid at position 109 of the neighboring HA monomer via a hydrogen bond. Moreover, amino acid 111 is as well dependent on the HA group, in group 1 it is almost always histidine and in group 2 it is almost always threonine. When the virus particle is acidified the amino acids 106 and 111 of HA2 are no longer part of the helix, but change into a loop [33].

Next to the conserved amino acids at position 51, 109 and 112, which play an important role in determining the fusion pH, cleavage of HA0 activates structurally some reorganization in HA2. Firstly, the coiled coil becomes longer, since a loop and α-helix are reinforcing the coilded coil: it becomes 38 amino acids longer and hereby shifting the fusion peptide, see red part in **figure 9** [48, 49]. Secondly, the α-helix in the middle of HA2 will unfold; green part in **figure 9**. Thirdly, residues 141 till 175 of HA2 (from the C terminal part) will unfold and align along the coiled coil; pink part in **figure 9** [22].

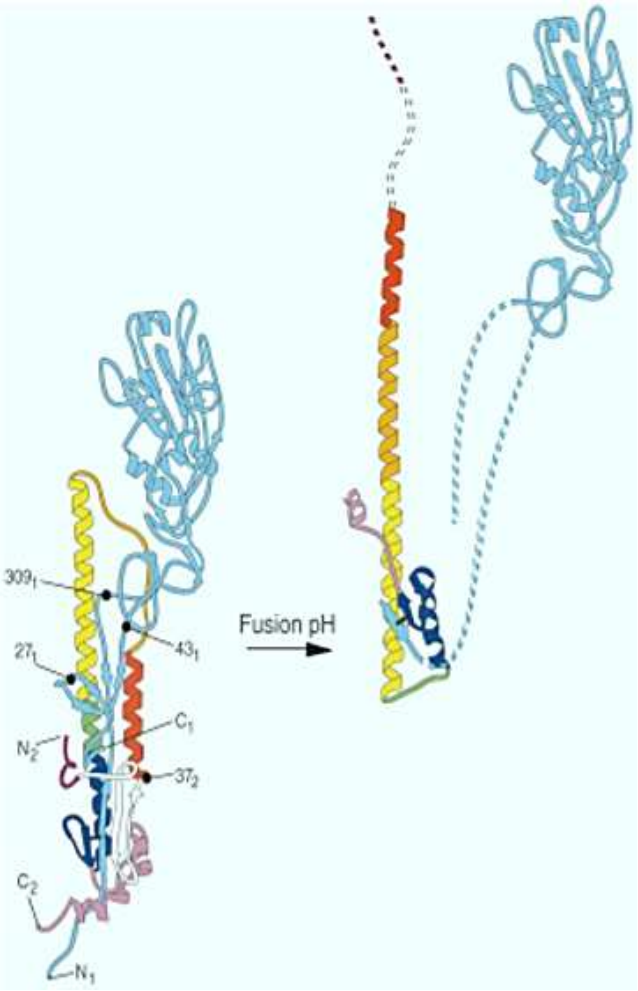

Figure 9. HA in neutral pH (left) and HA in fusion pH (right): the coiled coil is extended and an α-helix unfolds [22].

Looking at vaccines, most influenza antigens are based on HA (and NA). However, this is not completely correct, since the antigenic sites are solely located in HA1 [31, 50]. HA of H1 consists of four conserved antigenic sites, named: Ca, Cb, Sa and Sb. Ca consist of two sites (Ca1 and Ca2) and contacts two nearby receptor binding domains in the trimer, see **figure 10.** In addition, Sa is well conserved in the H1 pandemic viral strains of 1918 and 2009, and is not glycosylated [51]. HA of H3 consist of five conserved antigenic sites: A, B, C, D and E. Where antigenic site A is located at a loop, B at the tip, C at the hinge and D can be found in the inner side of the HA trimer [52]. Stray & Pittman [53] assigned the amino acids of the antigenic sites in their research for H1 as well as H3, see **tables 1** and **2** [53].

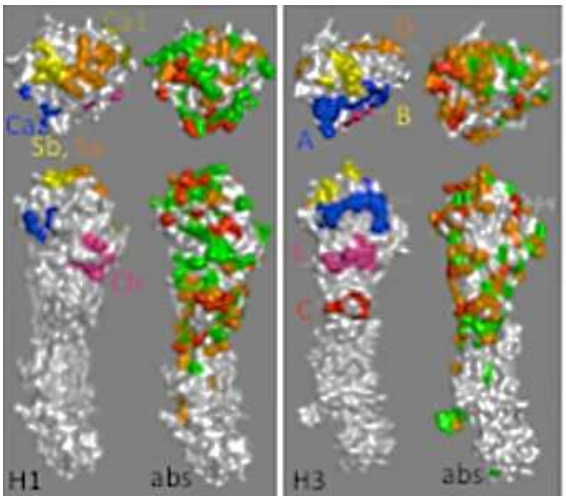

Figure 10. At the left side the antigenic sites of HA in H1 are shown. The upper figure is the top view. The lower figure is the side view. Olive is antigenic site Ca1, blue is Ca2, pink is Cb, orange is Sa and yellow is Sb. At the right side the antigenic sites of HA in H3 are shown. The upper figure is the top view. Blue is antigenic site A, yellow is B, red is C, orange is D and pink is E. The abs figures demonstrates the regions that are changed between different sequences, in which green is 50-75% of changed parts, orange is 75-90% and red is above 90% [53].

Table 1. Location of the antigenic sites in HA of H1 [53].

H1	Ca1	Ca2	Cb	Sa	Sb
Amino acids	169, 173, 207, 212, 240 and 242	140, 143-145, 149, 224 and 225	74, 75, 77-79 and 117	128, 129, 165-167	156, 159, 192, 193, 196 and 198

Table 2. Location of the antigenic sites in HA of H3 [53].

H3	A	B	C	D	E
Amino acids	122-127, 129, 132-138, 142-146	156-160, 186, 190, 193, 194, 196 and 197	50, 52-54, 275, 277 and 278	201-206, 217-220	62, 63, 79-83

Preliminary conclusion on the structure and functional domains of HA are that HA consist of seventeen subtypes, which can also be divided in highly pathogenic or low pathogenic depending on the amount of basic amino acids. Moreover, HA consists of two parts: the head and stem regions. The head domain function as the receptor binding pocket, were the type of amino acid determines whether HA is α(2, 3) – or α(2, 6) – linked to sialic acids. This influences which host organism can be infected. While the stem region, which is a well conserved region, is responsible for membrane fusion. During membrane fusion certain irreversible structural changes take place.

Location of glycosylation sites and length of the oligosaccharides

Glycosylation is a process accomplished by the Endoplasmic Reticulum (ER). In this process sugar groups, also referred to as glycans, are added to the protein. There are many types of glycans, that each have different lengths and different amount of side chains, it makes glycoproteins the ideal candidates for cell recognition, since they are very precise. Most oligosaccharides are N-linked, which means that they are bound to the NH_2 – group of an asparagine (Asn) residue in the protein, see **figure 11** [19, 54-57]. However, linkage is solely possible if the Asn residue has the amino sequence Asn-X-Ser or Asn-X-Thr in which Ser is serine, Thr stands for Threonine and X is any amino acid except

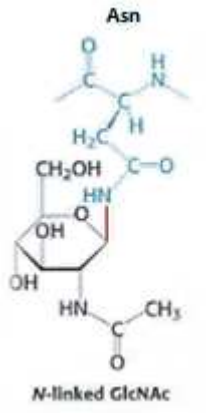

Figure 11. N-linked glycosylation. In blue is the asparagine residue. In red the glycosylation bond and in black is an oligosaccharide [55].

for proline [55, 57]. Further, linkage is aided by a precursor oligosaccharide which is located in the membrane of the ER. This precursor oligosaccharide is formatted piece by piece and then transferred to the protein that has to be glycosylated [54]. Nonetheless, there is a basic structure of the N-linked oligosaccharides: a penta-saccharide containing three mannoses and two N-acetylglucosamines. In addition, oligosaccharides are added to this basic structure. This helps to fold the protein properly [55], because when a protein is not properly folded it is not able to leave the ER. From the ER, the glycosylated protein is directed towards the Golgi complex where some oligosaccharides are trimmed. Due to this trimming the same protein can have a high diversity in forms [54, 55].

Besides the high diversity, glycosylation is as well necessary to creat infective virions [58]. Moreover, the glycans at the stem region are well conserved and are important for correct folding of HA. Additionally, they are required for the fusion process [32, 44, 57, 59-62], while the amount and structure can differ of the glycans in the head region of HA [50, 56, 57, 61, 63]. Investigations on the influence of receptor binding and immune response of the oligosaccharides that are present on HA have provided more insight in the binding abilities of HA. HA contains solely N-linked glycans. They are necessary for the virus particle to enter the host cell and regulate the structure and with that the functionality of HA [62, 64]. In addition, the structure of HA depends on conserved glycosylation sites and so affecting the fusion activity [44, 59].

Additionally, the function of conserved glycosylation spots in HA was analyzed by Roberts et al. [59] and Ohuchi et al. [44]. There are three conserved N-linked glycans in HA that can be found in multiple influenza A subtypes at position 12, 28 and 478 in the amino acid sequence. Mutagenesis experiments showed that the efficiency of transport of HA depends on these conserved N-linked glycans. Furthermore, the researchers made an extra form of HA in which an extra glycosylation site at position 482 was added, in this case the mutant lacked the glycosylation site at position 478, see **figure 12** for an overview. It was demonstrated that the location of the glycosylation sites is important for transport. When all three conserved glycosylation sites are deleted, HA is stuck in the ER. However, when only the glycosylation sites at positions 12 and 28 are deleted, HA is stuck in the Golgi system, which means that the conserved glycosylation site at position 478 is necessary to transport HA from the Golgi system to the plasma membrane. When two of the three conserved glycosylation sites were present, the transport of HA was functioning normal. In more detail; the conserved sites at position 12 and 28 are important. When one of these two glycosylation sites is present the transport of HA still functions normally, solely at a slower rate when compared to the transport in the wild type. Moreover, when the glycosylation site at position 12 is removed and when all three of the conserved glycosylation sites are deleted, it arises that there is a problem in the cleavage of HA. This might be due to the problem that the enzymes are not able to reach the

cleavage site or it might be that the transport of HA is inhibited due to incorrect folding [44, 59]. So the location of these conserved glycosylation sites is important for the cleavage of HA, the correct folding of HA and the transport.

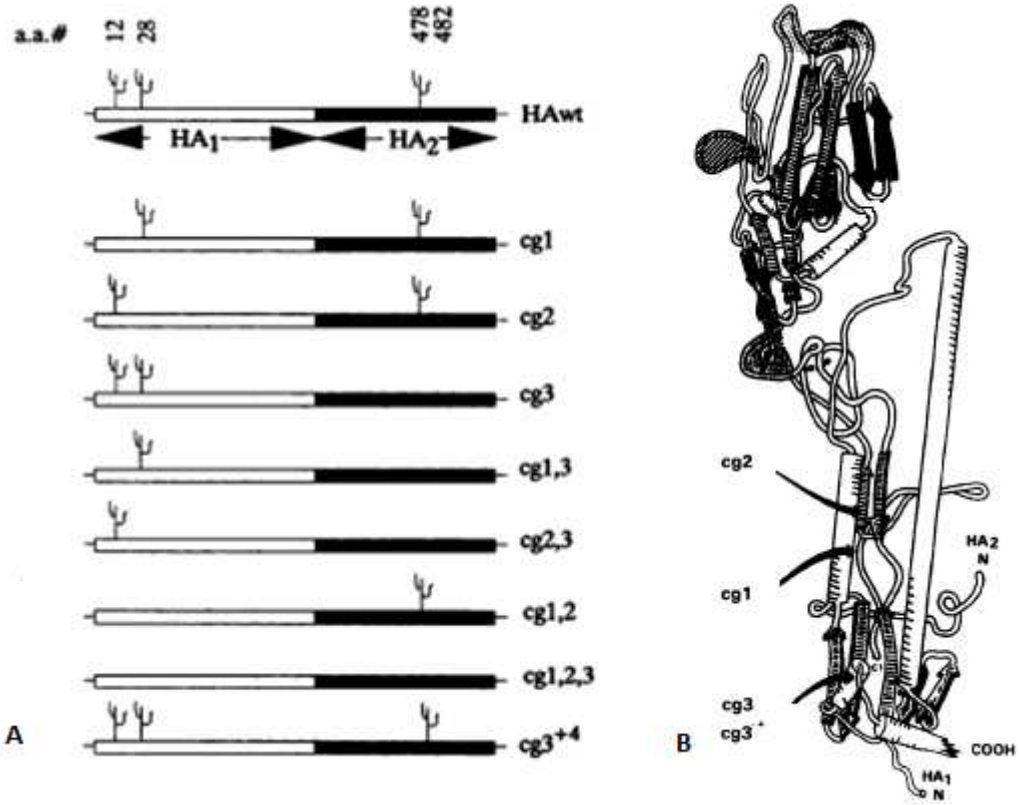

Figure 12. Structural representation of different glycosylated states which are conserved in the research of Roberts et al. [59] where certain glycans were deleted. A) Schematic overview of the conserved glycosylation (cg) sites. HAwt is the wild type strain. The figures on top of the HAwt are glycosylation sites, located by their amino acid number (a.a.#). B) Location of the conserved glycosylation sites in the three dimensional structure [59].

The amount of glycosylation sites in the head region of HA in H1 influenza has increased compared to the H1 strain of the Spanish flu, however they differ between viral strains [57]. The HA protein of the Spanish flu had only one N-glycosylated site positioned at residue 94 (H3 numbering) [50, 65, 66]. This strain has been inactive for some while until it reappeared in 1977, when it suddenly had five glycosylation sites at residues 94, 131, 158, 163 and 271 [57, 65]. Nowadays however, the seasonal H1 strain contains about four glycosylation sites at positions 63, 94, 129 and 163. It is thought that H1 strains are less susceptible for obtaining N-glycosylation sites compared to the H3 strain [57, 65]. Although it is suggested that H1 can have up to seven glycosylation sites [50, 65, 66]. These seven glycosylation sites are suggested at position 20, 21, 33, 63, 94, 130 and 163. Residue 163 is located in the antigenic site Sa. Furthermore, Zaraket et al. [67] suggest that there are four more possible glycosylation sites at position 144, 149, 163 and 190 [67]. However, when N-linked glycans are created in some regions this might hinder trafficking in the cell by diminishing the stability of the structure as

well the function may be disabled [66]. The extra possible glycosylation sites might suggest that in the future new glycosylated forms of HA might be found and thus new epidemics or even pandemics might occur.

The amount of glycosylation sites in the head region of HA in H3 influenza has increased over the last four decades to avoid the immune system. Although the pandemic Hong Kong flu of 1968 had only two N-glycosylation sites, nowadays the seasonal H3 group can contain up to eleven N-glycosylated sites in the head region of HA [65, 68]. The first N-glycosylation sites that were found next to the glycosylation sites of the Hong Kong strain are located at residues 63, 126 and 246 [65, 66]. Apparently, the extra glycosylation sites were needed to avoid immunity of the host that was infected with H3 strains that had less glycosylation sites. Later the glycosylation sites at positions 45, 122, 133 and 144 joined [65, 67]. From which residue 122, 126, 133 and 144 can be found in antigenic site A (**table 2**) [66, 67]. Additionally, residue 63 can be found in antigenic site E (**table 2**). The addition of an glycosylation site was accomplished by single and multiple amino acid substitutions [65]. Other glycosylation sites are suggested at the positions 22, 38, 122, 165, and 285 [65, 68, 69]. In addition, Zaraket et al. [67] demonstrated that the HA of H3N2 can also acquire another three possible glycosylation sites at position 50, 131 and 220 [67]. From which residue 220 can be found in antigenic site D (**table 2**).

Abe et al. [63] investigated the evolution of H3N2. Over the past three decades the HA head region received more glycosylation sites. HA that contains more oligosaccharide side chains appeared to be transported to the plasma membrane like the wild type. Although, when more glycosylation sites were present (three to six) the binding of the receptor was impaired. However, the virus was still active enough to survive; looking at the cell fusion in **table 3**[a]. When four or more glycans are present, the receptor binding of HA is weakened (compared to the wild type). When only one glycan is present, located on amino acid residue 165 (sample 1 in **table 3**), the transport and biological functions of HA appear to function normal. Apparently, this one oligosaccharide side chain at position 165 is necessary for H3 viruses to survive [63]. Furthermore, when the amount of possible glycosylation sites in the head region of HA increases in H3N2, the virulence of this virus decreases [57, 70].

The H3 group strains found at this moment can have a maximum of eleven possible glycosylation sites in the head region of HA, compared to the H1 group which can have up to five glycosylation sites [29, 41, 56, 63, 71]. From these five glycosylation sites of H1 a maximum of three can have an oligosaccharide side chain [41]. In addition, these glycosylation sites appeared to be conserved in all virus strains [29].

[a] Table 3 is derived from [63].

Table 3. Overview of HA transport and function in different glycosylation states [63].

HA	Amount of glycosylation sites	Oligosaccharide amount	Cell fusion	Cell surface expression	Guinea pig erythrocytes binding (% compared to wild type)	Chicken erythrocytes binding (% compared to wild type)
Wild type	2	2	+	+	100	100
1	1	1	+	+	111	115
2	2	2	+	+	110	99
3	3	3	+	+	97	36
4	4	4	+	+	25	8
5	5	4	+	+	17	4
6	5	5	+	+	16	6
7	6	5	+	+	11	3

Avian H5 is sporadically infecting humans, therefore research is required on this influenza strain before it might cause seasonal epidemics in humans. HA of avian H5 can have up to seven multi-basic amino acids in the cleavage site motives and nowadays H5 can contain up to six glycosylation sites depending on the viral strain. Four of these six glycosylation sites are located at residue 131, 158, 169 and 240 (H3 numbering) [72, 73]. The other two glycosylation sites are present at residues 15 and 27 according to Chen et al.[74]. However, until this point research on influenza HA has been focusing mostly on H1 and H3 subtypes.

Analysis demonstrated that the oligosaccharide side chains in HA1 of H1N1 consist of three different sorts of glycans: mannose-rich side chains, complex and these two types combined. The complex side chain contains galactose, mannose, fucose, and glucosamine. Their ratios are 6:11:2:5. It is thought, due to the high quantity of fucose and galactose, that the strains that were tested have three or four substantial carbohydrate side chains and only a low percentage that forms mannose-rich chains[58, 75]. Collins and Knight [58] found that fucose and galactose are located at the terminal part and sialic acids bind to the galactose, see **figure 5** [39, 58].

Schulze et al. [66] found that one glycosylation site is hidden between monomers of the HA trimer, which exists only of oligomannose glycans. The other glycosylation sites in HA, like the carbohydrate side chains that were found in H1N1, consist of multiple sugars with different sizes; the size of the glycans become smaller when they are closer to the lipid bilayer. Depending on the cell type where HA is grown in tri- or tetra glycans are formed [66, 75].

Preliminary conclusion on some of the locations of the glycosylation sites in HA, is that certain glycosylation sites are well conserved in all HA subtypes. These conserved sites are located at positions 12, 28 and 478. In addition, they control transport of the virus. More glycosylation sites are located at the receptor binding domain, however each viral strain has a different amount of glycosylated sites. This can go up to five in the H1 strains and up to eleven in the H3 strains. For H1 the glycosylated sites are spotted nowadays at residues 63, 94, 129 and 163 (H3 numbering). While for H3 at the moment the glycosylated sites are located at residues 45, 63, 122, 126, 133, 144 and 246. Furthermore, there are also some glycosylation sites at the stem region, however they are well conserved and do not tend to change as often like the oligosaccharide side chains in the head region. The length of the carbohydrate side chains in the head region varies for each glycan. It is thought that there are some long glycans which contain fucose and galactose. Furthermore, it is thought that only a small percentage forms mannose chains and another sugar that is present is glucosamine.

The influence of the glycosylation sites on the protein's functions

The glycosylation sites in the head region can be very diverse and certain glycans might disturb the recognition of antibodies by blocking the antibody binding site [18, 62]. To give some more insight: the receptor binding site is about 800 $Å^2$, while the area that is required for an antibody to bind is between 1200 and 1500 $Å^2$. This suggests that antibodies do not fit the receptor binding pocket [15]. However, somehow antibodies are able to bind this small binding pocket. Though, mutations around the receptor binding domain might interfere with the position of glycosylation which can disturb the binding of HA to sialic acid, for example by steric hindrance of the binding site. Furthermore, no internalization is measured when the glycosylated sites are deleted [64]. So the carbohydrate side chains are important for infectivity and they can mislead the immune system by masking and blocking the antibody binding sites [65, 67-69].

Researchers performed multiple experiments on HA and its glycans; some tested if the expression system is important, others check the location of the glycosylated sites [18, 59]. Roberts et al. [59] demonstrated that the amount of oligosaccharides and their location determine the rate of transport

and the shape of the protein and thereby the stability of the protein [59]. Furthermore, the receptor binding properties of HA trimers were studied in different expression systems to check whether the expression systems resulted in different kind of outcomes in the glycosylation of HA, since large and bulky side chains might cause steric hindrance to the binding pocket of the receptor. In addition, the side chains more or less determine the spot where HA is able to bind. De Vries et al. [18] experimented with different cell lines, insect cells and two types of human cell lines (HEK293S GnTI (-) and HEK293T) to investigate if the expression systems affect the binding affinity of HA, due to the different ways of glycosylation. They revealed that when the oligosaccharide side chains of HA becomes smaller, the specificity for binding to the receptor decreases as well. In this case that means that the insect cells produced the HA with the smallest oligosaccharide side chains, which is binding less specific to the receptor of HA and has an affinity for multiple sialic acid receptor types. So, while HA of insect cells have a high affinity, they also have a lower specificity [18].

Moreover, Ohuchi et al. [44] investigated the deletion of the glycosylation sites in the head region of HA, however this did not interfere with the fusion activity, which suggests that solely the conserved glycosylation sites in the stem region of HA are important for membrane fusion [44]. This caused the researchers to conclude that the role of these conserved glycans in HA are necessary to maintain the structure of HA in which it is transported, so correct folding depends on these conserved glycosylation sites in the stem region and with that the fusion structure and fusion activity [44, 59].

Likewise, Wang et al. [62] found that the peptide sequences near the glycosylation sites are well conserved and they analyzed these conserved spots. In order to generate different glycosylated proteins, HAs were produced in different cell lines and treated enzymatically. The different structures that are produced by the different cell lines can be found in **figure 13** [62]. The researchers found that all these glycosylation forms of HA prefer the sialic acid which is linked with an α(2, 3) – bond. This might be due to steric hindrance of the glycans for the receptor binding site. Additionally, mono-glycosylated HA (only containing GlcNAc) binds with the greatest affinity, however this HA does not have a specific type of receptor which it prefers. The researchers concluded that HA binds best to sialic acid when it is most simply structured: mono-glycosylated [62].

From these experiments it can be concluded that HA has a high affinity when there are less (truncated) oligosaccharides present. For example, when HA is compared in mammalian cells (high amount of glycosylation) and insect cells (low amount of glycosylation) [18, 62]. Moreover, Wang et al. [62] found that when HA contains short oligosaccharides it prefers an α(2, 3)– linkage, however this HA is then less specific. This makes an avian infection more susceptible [62].

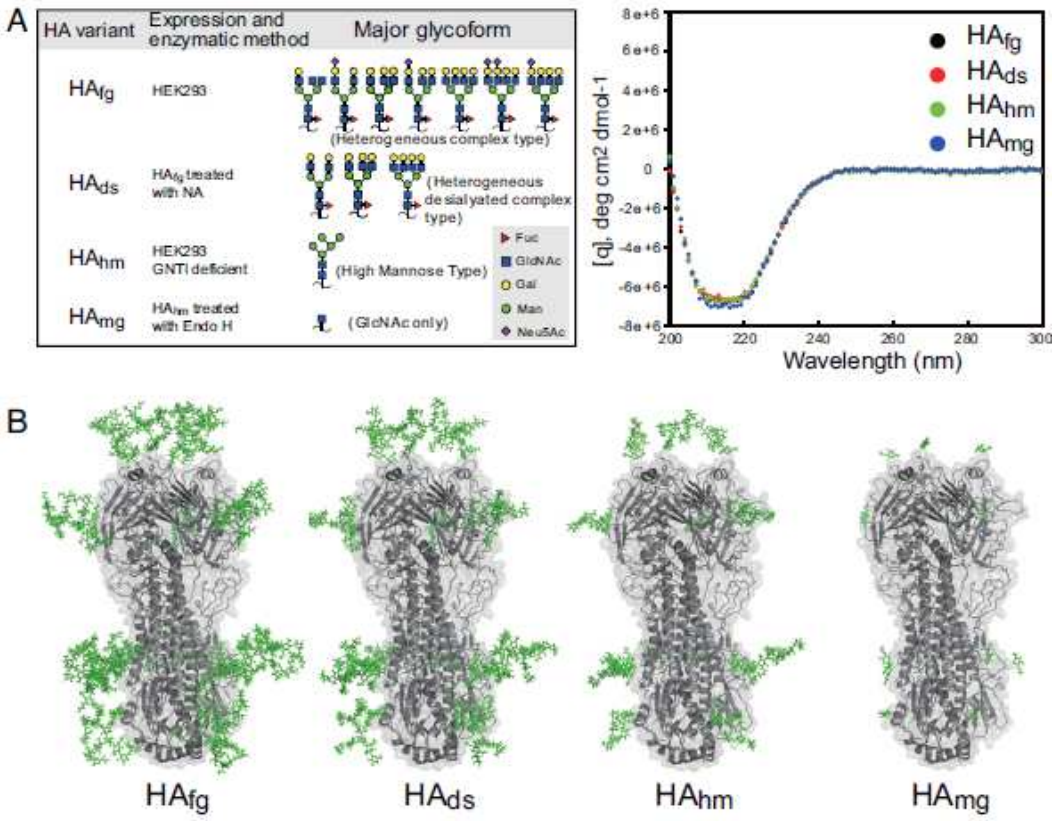

Figure 13. The different glycosylation forms of HA. HA_{fg}: completely glycosylated HA derived from HEK293S cells. HA_{ds}: mannose-rich HA derived from HEK293S cells which lack N-acetylglycosaminyl tranferase I (HEK293S GnTI(-)). HA_{hm}: HA of HEK293S cells that are treated with neuraminidase, which is an enzyme that cuts the sialic acid bond. HA_{mg}: mono-glycosylated HA of HEK293S GnTI(-) cells that are treated with Endo H, which is an enzyme that removes the glycans and leaves only one GlcNAc (which is a sugar) at the HA. (A) left; A schematic overview of the different glycosylation forms of HA. (A) right; Circular Dichroism of the four different HA structures. All lines overlap which means that all different HA forms have the same secondary structure. (B) Structures of the four different HA forms. In grey the protein structure is visible. In green the N-linked glycans are shown [62].

Preliminary conclusion on the influence of the glycosylation sites on the functions of HA is that the glycosylation sites in the stem region are important for the membrane fusion process and with that the correct folding of HA and the fusion activity. Glycosylation sites in the head region are important for receptor binding as well as masking the receptor binding domain from antibodies. Furthermore, glycans control HA cleavage and with that the infectivity of the virus.

Vaccine development

Depending on the virus, the response of the immune system and geographical location, the type of vaccine is chosen. For influenza viruses there are different types of vaccines available: live and attenuated vaccines, inactivated vaccines, subunit vaccines and DNA vaccines [76].

Live and attenuated vaccines are vaccines which have a weakened type of virus. This type of vaccine activates the immune system the same way as when the body normally is infected with a virus. The advantage of this weakened virus is that it might provide immunity for life. However, a disadvantage of this type of vaccine is that viruses can reverse back to a virulent form by secondary mutations and might cause severe illness. Furthermore, this type of vaccine has to be stored coldly, which might be a problem in third world countries. In addition, people who are ill or are immuno-compromised cannot have this type of vaccine [76].

Inactivated vaccines contain virus which was killed first in a laboratory. One advantage of this type of vaccines is that the viruses cannot mutate and is therefore stable. The vaccine also does not need to be stored at a cold temperature. A disadvantage on the other hand of this type of vaccines is that they cause a weaker immune response than live and attenuated vaccines, so probably more vaccine doses are required to acquire immunity for a longer time, which might be a problem for the people who live in third world countries since they do not always have access to the health system [76].

Subunit vaccines are vaccines that contain parts of the virus. They mostly only contain the part of the virus that activates the immune system: epitopes [76]. An epitope is an antigenic determinant which is recognized by a specific part of the antibody [77, 78]. Epitopes are (mostly) found at antigenic sites. However, for this type of vaccines research is necessary to discover the exact antigen that is activating the immune system, and this might cost a lot of time [76].

DNA vaccines are another option. An advantage of this type of vaccines is that the cells are making their own antigens. Furthermore, this type of vaccine cannot cause illness, since it does not contain the virus; only some genes of the virus are present. In addition, this type of vaccine is easy to produce at a low cost [76].

There are multiple strains of influenza which evolve each year. To prevent influenza pandemics the vaccine against influenza therefore has to be modified annually and has to be provided year by year to the population [2, 18, 29]. The flu vaccines that are provided each year are developed on the basis of the influenza virus strain of the previous year [79]. So each year a new vaccine has to be developed and each year certain risk groups (under which old people, people with asthma, young children, etcetera)

of the population has to be vaccinated, which is quit time consuming in general and of course research and development costs money. A solution for the problem of time and money would be to develop a universal vaccine.

The influenza virus evolves to preserve the ability of host infection, also called antigenic drift and shift [28, 29, 41, 56, 57, 80]. Antigenic drift is a change in the genotype which is caused by the immunity of the host; with antigenic drift the virus mutates to prevent the host from becoming immune [28]. In addition, antigenic shift is the reassortment of gene segments of different viral strains which can lead to new viral strains; this can occur when two different virus strains infect one host organism and the viral components of both viral strains are mixed into a newly formed virus particle [28, 81]. For example the two pandemics of 1957 and 1968 had a human as well as an avian source. Most probably due to the avian input these strains became this fatal. In another case, in 1976 in New Jersey, a strain with a pig as source was deathly for some patients, however pigs might not be the only carriers for influenza strains [82]. Furthermore, human strains can infect pigs, which is what happened in Italy with the pandemic strain of 2009, a H1N1 strain. In this case only the NA gene was not originated of the pandemic strain of 2009, which might be dangerous in the future. Since the virus can adapt more due to reassortment and when it then hits the human population again it might be even stronger than the pandemic of 2009 [83]. Moreover, for viral strains to be able to infect through air a human HA and polymerase subunit PB2 are necessary in avian influenza. If this is not the case the viral strain cannot be passed on through the air [84]. Due to these evolutional steps in influenza A viruses, research is required for the development of a universal vaccine. Analysis of this influenza evolution focuses on single mutations and the behavior of genes. Behavior analysis of genes can focus on unified changes between genes, like an amino acid substitution outside the epitope that has to restore the function of the virus caused by an amino acid substitution in the epitope [29, 41, 56, 57, 63, 80, 81].

Another adaptation in the evolution of the influenza virus is addition or deletion in the amount of glycosylation sites. Furthermore, the length of the glycans is important as well to cause antigenic drift [29, 44, 56, 57]. A positive effect of glycans at a possible glycosylation site can be that HA might bind less strong due to steric hindrance that the oligosaccharides give, so the receptor binding is weakened. This leads to easier viral release when new virus particles are formed. For example, when HA is produced in insect cells [18]. On the other hand, long oligosaccharide side chains can hinder antibody recognition due to coverage of the antibody binding site. Moreover, these oligosaccharides may have an effect on the cleavage of HA; in such a case it thus all depends on the location of the glycans [41, 44, 57, 61, 63, 70].

HA1 is continually under natural selection, to prevent the host cell from getting immune for HA. HA1 sequences dating from 1968 till 2005 were analyzed and it can be concluded that the viral strains undergo positive selection, which means that the amino acids that are changed in the sequences are not randomly. They are specifically picked, since almost all amino acid substitutions were located at the antigenic sites and in a certain time these substitutions are getting fixed. Most of these substitutions occur at the same time: mostly there is not one amino acid substitution, but multiple substitutions simultaneously. Certain amino acids undergo more than one substitution like the residues at position 145 and 156. On the other hand, there are also residues where relatively no substitutions take place, e.g. residue 138. Furthermore, residues 83, 155, 172 and 189 are positively selected to mutate. They are located in an antigenic site and underwent multiple substitutions. In the period that was studied they found that most mutations were made in an antigenic site, which is ideal to prevent immunity of the host organisms antibodies. Moreover, residues at positions 67, 78, 144 and 213 mutated and were fixed in HA [31].

The evolution of HA1 of H1N1 was as well studied by Zakaret, et al. [67]. The researchers found that 36 amino acid substitutions evolved in the studied time (which is approximately ten years). From these 36 mutations solely eleven were present for more than one year and some of these mutations can be found in the receptor binding pocket and antigenic sites (positions 190 (in the 190-helix), 124 (in the antigenic site Sa), 169 and 270 (in the antigenic site Ca1; **table 1**)) [67].

Although the H1 group is evolving at a slower rate than the H3 group, researchers found that in the viral strain that cost a pandemic in 2009, no antigenic drift had occurred: it was similar as the strain that caused the Spanish flu in 1918. This might be due to the fact that the virus was passed over to humans from pigs and apparently the H1N1 strain did not undergo antigenic drift in pigs which might be due to the life span of pigs [28, 57].

The ideal influenza vaccine is a vaccine that can bind to all influenza strains and at the same time it gives the host immunity for a longer amount of time, so it has to activate the immune system. At this moment the focus for new vaccines is on the conserved parts of HA to derive antigens that target conserved epitopes in all HAs. Antibodies against the stem region might be the solution, since the stem region is conserved in all subtypes of influenza A HAs. Actually, the head region is the part which activates the immune response of the host cell and it is also the region were receptor binding takes place, however this is too specific for all the different viral strains [41]. By blocking the epitopes in the head region through long glycans it might be possible to direct the antibody response to the stem region instead of the head region. This way it may be possible to derive an antibody response based on the stem region with all its conserved glycans which would give universal antibodies against the

influenza virus. At this moment there are already monoclonal antibodies found against the stem region of HA. Multiple research groups worked on universal monoclonal antibodies against the influenza virus. Some research groups derived the monoclonal antibodies by phage display selections that are synthetically made, others from someone who got vaccinated [85, 86]. One of the groups focused on antibodies that were derived from a vaccinated person, this vaccination was against the H1N1 strain of 2009. Some of the antibodies that were derived by this group, are effective against different viral strains of the influenza virus [85]. Another group tested their monoclonal antibodies on mice that were exposed to different influenza strains. They concluded that the mice in these tests were protected against influenza strains which are group 1 classified (the H1 group) and not against influenza strains from group 2. The researchers of that groups suggests that a glycosylation site in group 2 classified HA1 is responsible for this [86].

Preliminary conclusion on the vaccination programs that are present at the moment is that there are four different vaccines possible. All of them have their own advantages and disadvantages and it depends on the situation which one should be used. The annual vaccination programs are necessary since the influenza strains are evolving in a high speed so that each year a vaccination is required for the weaker groups in the population. A universal vaccine would be the ideal solution against the different influenza viral strains, which could be acquired by obtaining antibodies against the stem region of HA. Since this is the part of HA which has the most conserved immunogenic domain in all HA subtypes. Nowadays, there are already monoclonal antibodies derived by research groups against the stem region. Some of these antibodies are effective against different influenza subtypes, others are only working against group 1 classified HA.

Conclusion

In summary, vaccines are developed against influenza. They are based on the antigenic properties of HA and NA of the viral strain that was circulating the previous year. Therefore each year a new vaccine has to be developed which costs research, development and annually the population needs to be vaccinated. However, HA can have different conformations due to glycosylation, the amount of glycosylation sites can differ as well as the length of the glycans. This is influenced by the expression system that is used. In addition, the affinity and specificity of HA depends on the amount of oligosaccharides that are present. Furthermore, conserved glycosylation sites in the stem region are important for the correct folding of HA for transport. This should all be taken into account when a vaccine against HA is produced. A universal vaccine would be the ideal solution for the annual

vaccinations, which should focus on the stem region of HA, since this part of HA is well conserved in all HA subtypes and so the same vaccine can be used each year against each viral strain that is causing an epidemic. Momentarily, there are monoclonal antibodies derived against HA, however they are not all as effective, since some do bind to multiple influenza subtypes, while others only bind to the group 1 HAs. So more research is required on antibodies against HA.

References

1. Nayak, D. P., Hui, E. K. & Barman, S. Assembly and budding of influenza virus. *Virus Res.* 106, 147-165 (2004).

2. Lamb R. & Krug, R. *Orthomyxoviridae*: the viruses and their replication. *Fields Virology*, 1487 - 1531 (2001).

3. Barry, J. M. & Noymer, A. The Great Influenza: The Epic Story of the Deadliest Plague in History. *Population and Development Review* 30, 537-539 (2004).

4. Palese, P. Influenza: old and new threats. *Nat. Med.* (2004).

5. Ghendon, Y. Introduction to Pandemic Influenza through History. *Eur. J. Epidemiol.* 10, 451-453 (1994).

6. Lamb, R. Death by influenza virus protein. *Nat. Med.* 7, 1286 (2001).

7. Omić, J. & Van Mieghem, P. Pandemics and Networks: the Case of the Mexican Flu. *Proceedings of the World Congress on Engineering and Computer Science* 2 (2010).

8. Chen, Y. *et al*. Human infections with the emerging avian influenza A H7N9 virus from wet market poultry: clinical analysis and characterisation of viral genome. *The Lancet* 381, 1916-1925 (2013).

9. Nayak, D. P., Balogun, R. A., Yamada, H., Zhou, Z. H. & Barman, S. Influenza virus morphogenesis and budding. *Virus Res.* 143, 147-161 (2009).

10. Noda, T. *et al*. Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature* 439, 490-492 (2006).

11. http://www.uniprot.org/uniprot/Q1I0S4visited on May 29, 2013.

12. Alberts, B. *et al*. in *Molecular biology of the cell* (eds Anderson, M. S. & Dilernia, B.) 1431-1432, 1433, 1439, 1440 (Garland Science, United States of America, 2002).

13. Subbarao, K. Scientific barriers to developing vaccines against avian influenza viruses. *Nature reviews.Immunology* 7, 267 (2007). Quote on page 7 is retrieved from this article: page 268.

14. Wu, H. *et al*. Integrity of a recombinant hemagglutinin protein of an avian influenza virus. *Biotechnol. Lett.* 31, 1511-1517 (2009).

15. Skehel, J. An overview of influenza haemagglutinin and neuraminidase. *Biologicals* 37, 177-178 (2009).

16. Russell, R. J. *et al*. Structure of Influenza Hemagglutinin in Complex with an Inhibitor of Membrane Fusion. *Proc. Natl. Acad. Sci. U. S. A.* 105, 17736-17741 (2008).

17. http://www.uniprot.org/uniprot/P03437 visited on April 3, 2013.

18. de Vries, R. P. *et al*. The influenza A virus hemagglutinin glycosylation state affects receptor-binding specificity. *Virology* 403, 17-25 (2010).

19. Farahmand, B. *et al*. Influenza Virus Hemagglutinin: A Model for Protein N-Glycosylation in Recombinant Escherichia coli. *Intervirology* 55, 219-224 (2012).

20. Chang, D. K., Cheng, S. F., Kantchev, E. A. B., Lin, C. H. & Liu, Y. T. Membrane interaction and structure of the transmembrane domain of influenza hemagglutinin and its fusion peptide complex. *BioMedCentral Biology* 6 (2008). Quote on page 8 is retrieved from this article: page 1.

21. Kasson, P. Receptor Binding by Influenza Virus: Using Computational Techniques To Extend Structural Data. *Biochemistry (N. Y.)* 51, 2359 (2012).

22. Skehel, J. J. & Wiley, D. C. RECEPTOR BINDING AND MEMBRANE FUSION IN VIRUS ENTRY: The Influenza Hemagglutinin. *Annu. Rev. Biochem.* 69, 531 (2000).

23. Hu, W. Identification of highly conserved domains in hemagglutinin associated with the receptor binding specificity of influenza viruses: 2009 H1N1, avian H5N1, and swine H1N2. *Journal Biomedical Science and Engineering* 3, 114-123 (2010).

24. Bullough, P. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* 371, 37 (1994).

25. Fouchier, R. A. M. *et al*. Characterization of a Novel Influenza A Virus Hemagglutinin Subtype (H16) Obtained from Black-Headed Gulls. *J. Virol.* 79, 2814-2822 (2005).

26. Tong, S. *et al*. A distinct lineage of influenza A virus from bats. - *Proceedings of the National Academy of Sciences*, 4269 (2012).

27. Russell, R. J., Stevens, D. J., Haire, L. F., Gamblin, S. J. & Skehel, J. J. Avian and human receptor binding by hemagglutinins of influenza A viruses. *Glycoconjugate Journal* 23, 85-92 (2006).

28. Medina, R. A. & García-Sastre, A. Influenza A viruses: new research developments. *Nature Reviews Microbiology* 9, 590-603 (2011). Quote on page 3 is retrieved from this article: page 590.

29. Zaraket, H. *et al*. Molecular evolution of human influenza A viruses in a local area
during eight influenza epidemics from 2000 to 2007. *Archives of Virology* 154, 285-295 (2009).

30. Hamilton, B. S., Whittaker, G. R. & Daniel, S. Influenza Virus-Mediated Membrane Fusion: Determinants of Hemagglutinin Fusogenic Activity and Experimental Approaches for Assessing Virus Fusion. *Viruses* 4, 1144-1168 (2012).

31. Shih, A. C., Hsiao, T., Ho, M. & Li, W. Simultaneous Amino Acid Substitutions at Antigenic Sites Drive Influenza A Hemagglutinin Evolution. *Proc. Natl. Acad. Sci. U. S. A.* 104, 6283-6288 (2007).

32. Nobusawa, E. *et al*. Comparison of complete amino acid sequences and receptor-binding properties among 13 serotypes of hemagglutinins of influenza A viruses. *Virology* 182, 475-485 (1991).

33. Thoennes, S. *et al*. Analysis of residues near the fusion peptide in the influenza hemagglutinin structure for roles in triggering membrane fusion. *Virology* 370, 403-414 (2008).

34. Luo, G. *et al*. Molecular mechanism underlying the action of a novel fusion inhibitor of influenza A virus.. *Journal of Virology* 71, 4062-4070 (1997).

35. Yen, H. L. *et al*. Changes in H5N1 influenza virus hemagglutinin receptor binding domain affect systemic spread. - *Proceedings of the National Academy of Sciences* 106, 286-291 (2009).

36. Wilson, I. A., Skehel, J. J. & Wiley, D. C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 angst resolution. *Nature* 289, 366-373 (1981).

37. Tumpey, T. M. *et al*. A Two-Amino Acid Change in the Hemagglutinin of the 1918 Influenza Virus Abolishes Transmission. *Science* 315, 655-659 (2007).

38. Kuiken, T. & Taubenberger, J. K. Pathology of human influenza revisited. *Vaccine* 26, Supplement 4, D59-D66 (2008).

39. Gamblin, S. J. & Skehel, J. J. Influenza Hemagglutinin and Neuraminidase Membrane Glycoproteins. *Journal of Biological Chemistry* 285, 28403-28409 (2010).

40. Martín, J. *et al*. Studies of the Binding Properties of Influenza Hemagglutinin Receptor-Site Mutants. *Virology* 241, 101-111 (1998).

41. Kim, J. I. & Park, M. S. N-Linked Glycosylation in the Hemagglutinin of Influenza A Viruses. *Yonsei Medical Journal* 53, 886-893 (2012).

42. Luo, G. *et al*. Molecular mechanism underlying the action of a novel fusion inhibitor of influenza A virus.. *Journal of Virology* 71, 4062-4070 (1997).

43. Xu, R., McBride, R., Nycholat, C. M., Paulson, J. C. & Wilson, I. A. Structural Characterization of the Hemagglutinin Receptor Specificity from the 2009 H1N1 Influenza Pandemic. *J. Virol.* 86, 982-990 (2012).

44. Ohuchi, R., Ohuchi, M., Garten, W. & Klenk, H. D. Oligosaccharides in the Stem Region Maintain the Influenza Virus Hemagglutinin in the Metastable Form Required for Fusion Activity. *Journal of Virology* 71, 3719-3725 (1997).

45. Jin, H., Leser, G. P. & Lamb, R. A. The influenza virus hemagglutinin cytoplasmic tail is not essential for virus assembly or infectivity. *The EMBO Journal* 13, 5504-5515 (1994).

46. Thoennes, S. *et al*. Analysis of residues near the fusion peptide in the influenza hemagglutinin structure for roles in triggering membrane fusion. *Virology* 370, 403-414 (2008).

47. Li, Z. *et al*. Length Requirements for Membrane Fusion of Influenza Virus Hemagglutinin Peptide Linkers to Transmembrane or Fusion Peptide Domains. - *Journal of Virology* 82, 6337-6348 (2008).

48. Chambers, P., Pringle, C. R. & Easton, A. J. Heptad repeat sequences are located adjacent to hydrofobic regions in several types of virus fusion glycoproteins. *Journal of General Virology* 71, 3075-3080 (1990).

49. Ward, C. W. & Dopheide, T. A. Influenza virus haemagglutinin. Structural predictions suggest that the fibrillar appearance is due to the presence of a coiled-coil.. *Australian Journal of Biological Science* 33, 441-447 (1980).

50. Basak, S., Pritchard, D. G., Bhown, A. S. & Compans, R. W. Glycosylation sites of influenza viral glycoproteins: characterization of tryptic glycopeptides from the A/USSR(H1N1) hemagglutinin glycoprotein.. *Journal of Virology* 37, 549-558 (1981).

51. DuBois, R. M. *et al*. The Receptor-Binding Domain of Influenza Virus Hemagglutinin Produced in Escherichia coli Folds into Its Native, Immunogenic Structure. - *Journal of Virology* 85, 865-872 (2011).

52. Jackson, D. C., Murray, J. M., White, D. O. & Gerhard, W. U. Enumeration of antigenic sites of influenza virus hemagglutinin.. *Infection and immunity* 37, 912-918 (1982).

53. Stray, S. J. & Pittman, L. B. Subtype- and antigenic site-specific differences in biophysical influences on evolution of influenza virus hemagglutinin. *Virology Journal* 9, 1-16 (2012).

54. Alberts, B. *et al*. in *Molecular Biology of the Cell* (eds Anderson, M. S. & Dilernia, B.) 612-613, 702-703 (Garland Science, United States of America, 2002).

55. Berg, J. M., Tymoczko, J. L. & Styer, L. in *Biochemistry* (ed Zimmerman, P. et al.) 316-320 (W. H. Freeman & Company, New York, 2007).

56. Tate, M. D., Job, E. R., Brooks, A. G. & Reading, P. C. Glycosylation of the hemagglutinin modulates the sensitivity of H3N2 influenza viruses to innate proteins in airway secretions and virulence in mice. *Virology* 413, 84-92 (2011).

57. Tate, M. D., Brooks, A. G. & Reading, P. C. Specific Sites of N-Linked Glycosylation on the Hemagglutinin of H1N1 Subtype Influenza A Virus Determine Sensitivity to Inhibitors of the Innate Immune System and Virulence in Mice. - *The Journal of Immunology* 187, 1884-1894 (2011).

58. Collins, J. K. & Knight, C. A. Purification of the influenza hemagglutinin glycoprotein and characterization of its carbohydrate components.. *Journal of Virology* 26, 457-467 (1978).

59. Roberts, P. C., Garten, W. & Klenk, H. D. Role of conserved glycosylation sites in maturation and transport of influenza A virus hemagglutinin. *Journal of Virology* 67, 3048-3060 (1993).

60. Daniels, R., Kurowski, B., Johnson, A. E. & Hebert, D. N. N-Linked Glycans Direct the Cotranslational Folding Pathway of Influenza Hemagglutinin. *Mol. Cell* 11, 79-90 (2003).

61. Wagner, R., Wolff, T., Herwig, A., Pleschka, S. & Klenk, H. D. Interdependence of Hemagglutinin Glycosylation and Neuraminidase as Regulators of Influenza Virus Growth: a Study by Reverse Genetics. *Journal of Virology* 74, 6316-6323 (2000).

62. Wang, C. *et al*. Glycans on influenza hemagglutinin affect receptor binding and immune response. *Proceedings of the National Academy of Sciences* 106, 18137-18142 (2009).

63. Abe, Y. *et al*. Effect of the Addition of Oligosaccharides on the Biological Activities and Antigenicity of Influenza A/H3N2 Virus Hemagglutinin. *Journal of Virology* 78, 9605-9611 (2004).

64. Chu, V. C. & Whittaker, G. R. Influenza Virus Entry and Infection Require Host Cell N-Linked Glycoprotein. *Proc. Natl. Acad. Sci. U. S. A.* 101, 18153-18158 (2004).

65. Igarashi, M., Ito, K., Kida, H. & Takada, A. Genetically destined potentials for N-linked glycosylation of influenza virus hemagglutinin. *Virology* 376, 323-329 (2008).

66. Schulze, I. T. Effects of Glycosylation on the Properties and Functions of Influenza Virus Hemagglutinin. *Journal of Infectious Diseases* 176, S24-S28 (1997).

67. Zaraket, H. *et al*. Molecular evolution of human influenza A viruses in a local area during eight influenza epidemics from 2000 to 2007. *Archives of Virology* 154, 285-295 (2009).

68. Zhang, M. *et al*. Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology* 14, 1229-1246 (2004).

69. Skehel, J. J. *et al*. A Carbohydrate Side Chain on Hemagglutinins of Hong Kong Influenza Viruses Inhibits Recognition by a Monoclonal Antibody. *Proc. Natl. Acad. Sci. U. S. A.* 81, 1779-1783 (1984).

70. Vigerust, D. J. *et al*. N-Linked Glycosylation Attenuates H3N2 Influenza Viruses. - *Journal of Virology*, 8593 (2007).

71. Cherry, J. L., Lipman, D. J., Nikolskaya, A. & Wolf, Y. I. Evolutionary Dynamics of N-Glycosylation Sites of Influenza Virus Hemagglutinin. *PLoS Current* 1, RRN1001. (2009).

72. Matrosovich, M., Zhou, N., Kawaoka, Y. & Webster, R. The Surface Glycoproteins of H5 Influenza Viruses Isolated from Humans, Chickens, and Wild Aquatic Birds Have Distinguishable Properties. *Journal of Virology* 73, 1146-1155 (1999).

73. Matsuoka, Y. *et al*. Neuraminidase Stalk Length and Additional Glycosylation of the Hemagglutinin Influence the Virulence of Influenza H5N1 Viruses for Mice. *Journal of Virology* 83, 4704-4708 (2009).

74. Chen, W., Sun, S. & Li, Z. Two Glycosylation Sites in H5N1 Influenza Virus Hemagglutinin That Affect Binding Preference by Computer-Based Analysis. *PLoS ONE* 7, 1-14 (2012).

75. Collins, J. K. & Knight, C. A. Removal of carbohydrate from influenza A virus and its hemagglutinin and the effect on biological activities.. *Journal of Virology* 27, 164-171 (1978).

76. http://www.niaid.nih.gov/topics/vaccines/understanding/pages/typesvaccines.aspx visited on June 26, 2013.

77. Alberts, B. *et al*. in *Molecular Biology of the cell* (eds Anderson, M. S. & Dilernia, B.) 1369-1370 (Garland Science, United States of America, 2002).

78. Berg, J. M., Tymoczko, J. L. & Stryer, L. in *Biochemistry* (ed Ahr, K. et al.) 84-85 (W.H. Freeman and company, New York, 2007).

79. Sandbulte, M. R. *et al*. Cross-Reactive Neuraminidase Antibodies Afford Partial Protection against H5N1 in Mice and Are Present in Unexposed Humans. *PLoS Medicine* 4, 0265-0272 (2007).

80. Du, X. *et al*. Networks of genomic co-occurrence capture characteristics of human influenza A (H3N2) evolution. *Genome research* 18, 178-187 (2008).

81. Bragstad, K., Nielsen, L. & Fomsgaard, A. The evolution of human influenza A viruses from 1999 to 2006: A complete genome study. *Virology Journal* 5, 40 (2008).

82. Shu, L. P. *et al*. Genetic Reassortment in Pandemic and Interpandemic Influenza Viruses: A Study of 122 Viruses Infecting Humans. *Eur. J. Epidemiol.* 12, 63-70 (1996).

83. Moreno, A. *et al*. Novel H1N2 swine influenza reassortant strain in pigs derived from the pandemic H1N1/2009 virus. *Vet. Microbiol.* 149, 472-477 (2011).

84. Van Hoeven, N. *et al*. Human HA and polymerase subunit PB2 proteins confer transmission of an avian influenza virus through the air. *Proceedings of the National Academy of Sciences* 106, 3366-3371 (2009).

85. Hu, W. *et al*. Fully human broadly neutralizing monoclonal antibodies against influenza A viruses generated from the memory B cells of a 2009 pandemic H1N1 influenza vaccine recipient. *Virology* 435, 320-328 (2013).

86. Corti, D. *et al*. Heterosubtypic neutralizing antibodies are produced by individuals immunized with a seasonal influenza vaccine. *The Journal of Clinical Investigation* 120, 1663-1673 (2010).