UNIVERSITA' DEGLI STUDI
DI MILANO – BICOCCA

SCUOLA DI DOTTORATO DI SCIENZE

Facoltà di Scienze Matematiche, Fisiche e Naturali

Corso di Dottorato di Ricerca in Biotecnologie Industriali XXVII ciclo

NMR as a tool for structural characterization of carbohydrates and glycan-protein interactions

Tutor: Prof. Barbara La Ferla
Co-Tutor: Dr. Marco Guerrini

Matr. 063552
Eleonora Macchi

Anno Accademico 2013/2014
Dipartimento di Biotecnologie e Bioscienze

Index

Introduction and General Background

1. Influenza

Influenza is a serious respiratory illness, which can be debiliting and cause complications that lead to hospitalisation and death.

The global burden of influenza epidemics, every year is believed to be 3-5 million cases of severe illness and 300000-500000 deaths. Pandemics, in contrast to epidemics, are rare events that occur every ten to fifty years, and their mortality impact ranges from devasting to moderate or mild.[2]

Human influenza viruses belong to the orthomyxovirus family, which consists of the following genera: influenza A, B and C virus, but only A and B viruses in humans are of epidemiological interest.

Influenza viruses are enveloped single-stranded RNA viruses with a pleomorphic appearance, and an average diameter of 120 nm.

The genomes of the influenza A and B viruses consist of 8 separate segments covered by the nucleocapsid protein. Togheter these build the ribonucleoprotein (RNP), each segment codes for a functionally important protein:

1. Polymerase B2 protein (PB2)
2. Polymerase B1 protein (PB1)
3. Polymerase A protein (PA)
4. Haemagglutinin (HA or H)
5. Nucleocapsid protein (NP)
6. Neuraminidase (NA or N)
7. Matrix protein (M): M1 constructs the matrix, and M2 acts as an ion channel pump to lower or maintain the pH of the endosome, only in influenza A virus

8. Non- structural protein (NS), with an hypothetical function

The proteins NS1 and NS2 present a regulatory function to promote the synthesis of viral components in the infected cell. The RNA-RNA polymerase is formed from PB2, PB1 and PA, and it is responsable for replication and transcriptation. It has an endonuclease activity and is linked to the RNP.

The envelope of the virus consists in a lipid bilayer membrane which originates from the virus producing cell and which contains prominent projections formed by HA and NA, as well as the M2 protein (Fig. 1). The matrix formed by the M1 protein is covered by the lipid layer.

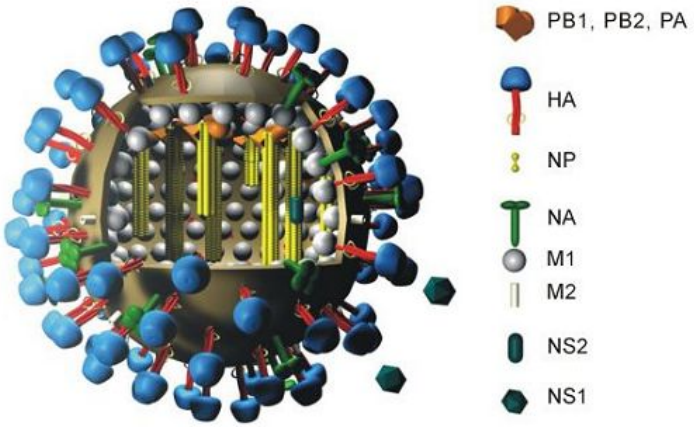


Fig. 1 Structure of an influenza A virus.

Influenza C virus has only 7 genome segments, and its surface carries only the glycoproteins moreover it has a low pathogenicity in humans.

Heamagglutinin (HA) is a glycoprotein containing either two or three glycosylation sites, with a molecular weight of approximately 76.000. The role of HA is to bind the sialic acid (N-acetyl-neurmainic acid, or Neu5Ac) presents in the host cells and induces penetration of the interior of the virus particle by

membrane fusion. Heamagglutinin is the main influenza antigen, and the antigenics sites are presented at the head of the molecule, while the feet are embedded in the lipid layer. The stalk region and the fusiogenic domain, which is needed for membrane fusion when the virus infects new cells, are contained in the body of the molecule. The HA forms trimers and several trimers form a fusion pore.

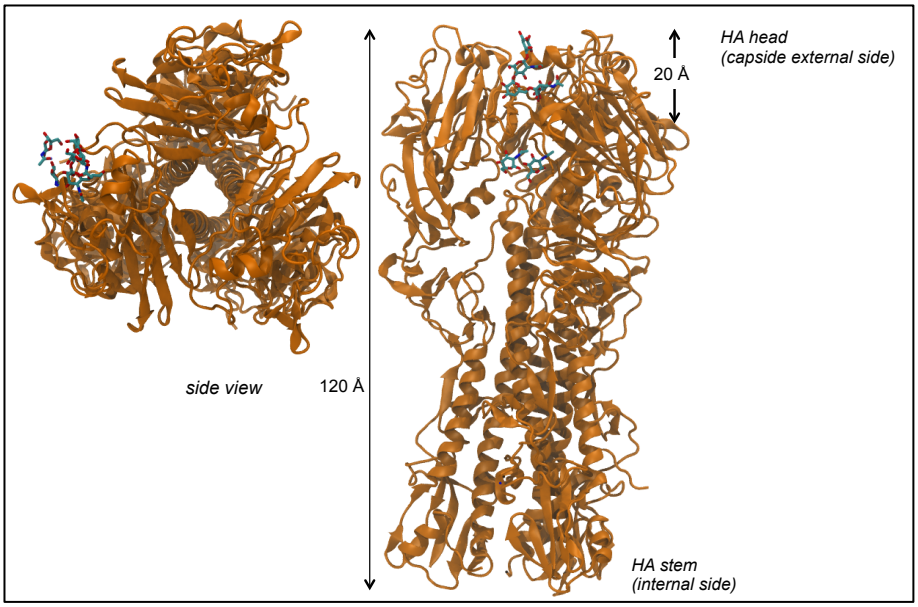


Fig. 2 HA structure

Every 2-8 years, in response to selection pressure to evade human immunity[3][4], there is the gradual evolution of viral strains, due to frequent mutations, called antigenic drift[5]. The antigenic drift process (Fig. 3) is subtle, involving point mutations within antibody-binding sites in the HA protein, and or NA protein, which potentially occur each time the virus replicates. The major part of these mutations are “neutral” as they do not affect the conformation of

the proteins; nevertheless, some mutations cause changes to the viral proteins affecting the binding of the host antibodies [6]. The mutations that cause the antigenic drift are the molecular explanation for the seasonal influenza epidemics during winter time in temperate climatic zones[7]; in fact the 'new infecting viruses' can no longer be inhibited effectively by host antibodies raised to previously circulating strains, allowing the virus to spread faster among the population[8].

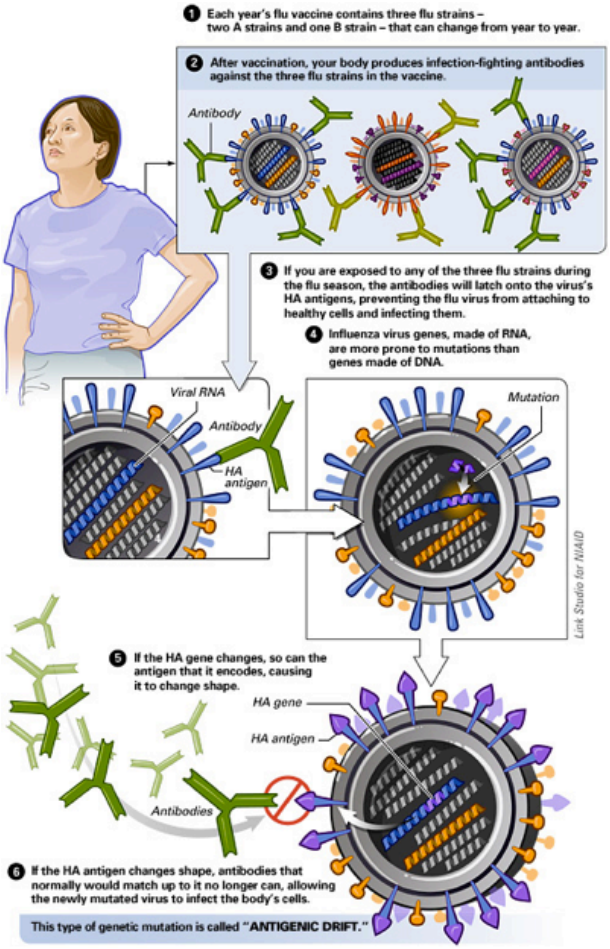


Fig. 3 Antigenic drift scheme [7]

The immune response is the basis for resolving the infection in an individual.

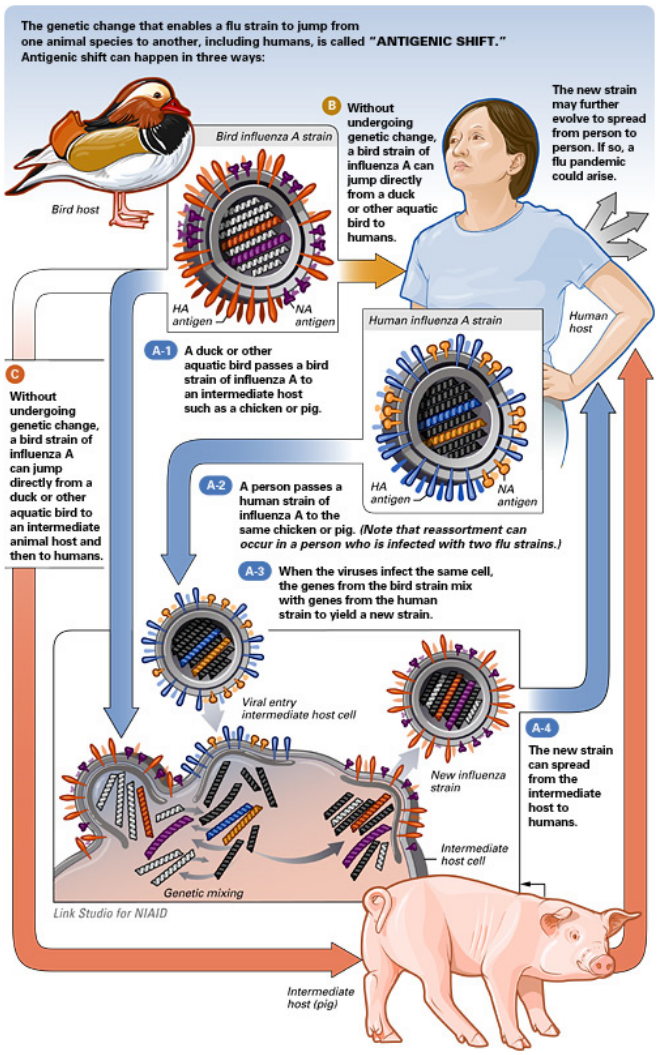


Fig. 4 Antigenic shift scheme [7]

Antigenic shift (Fig. 4), also called reassortment or genome reassortment, is only seen in influenza A viruses[6], and it arises when the HA is exchanged in a virus, for example H1 replaced by H5 resulting in a mosaic virus formation,

which have never been present in human circulation or last circulated decades before.

This shift may happen when a cell is infected by two different influenza viruses and their genome segments are exchanged during replication, and may have a significant impact causing pandemics or worldwide epidemics [9]. This genome reassortment is frequently seen in water birds, especially ducks. Once a virus has undergone antigenic shift, it remains susceptible to antigenic drift, as happens with any influenza virus. Moreover, all currently circulating influenza viruses are drift variants of previously pandemic influenza strains.

At this time, a major concern is the possibility that the highly pathogenic A/H5N1 avian influenza strain undergoes antigenic drift in such a manner that makes human-to-human transmissibility possible, resulting in a major worldwide human pandemic.

Like HA, neuraminidase (NA) is a glycoprotein present on the surface of the virus. It presents a tetrameric structure with an average molecular weight of 220.000. NA operates as an enzyme, and its role is to cleave sialic acid from the HA molecule, from other NA molecules and from glycoproteins and glycolipids at the cell surface. The NA molecule also acts as an important antigenic site, and in addition, seems to be necessary for the penetration of the virus through the mucin layer of the respiratory epithelium. Antigenic drift can also occur in the NA[7].

1.1 Replication cycle

The influenza virus life cycle can be divided into the following four foundamental steps: (1) attachment and entry, (2) fusion and uncoating, (3) transcription and replication and (4) assembly and release.

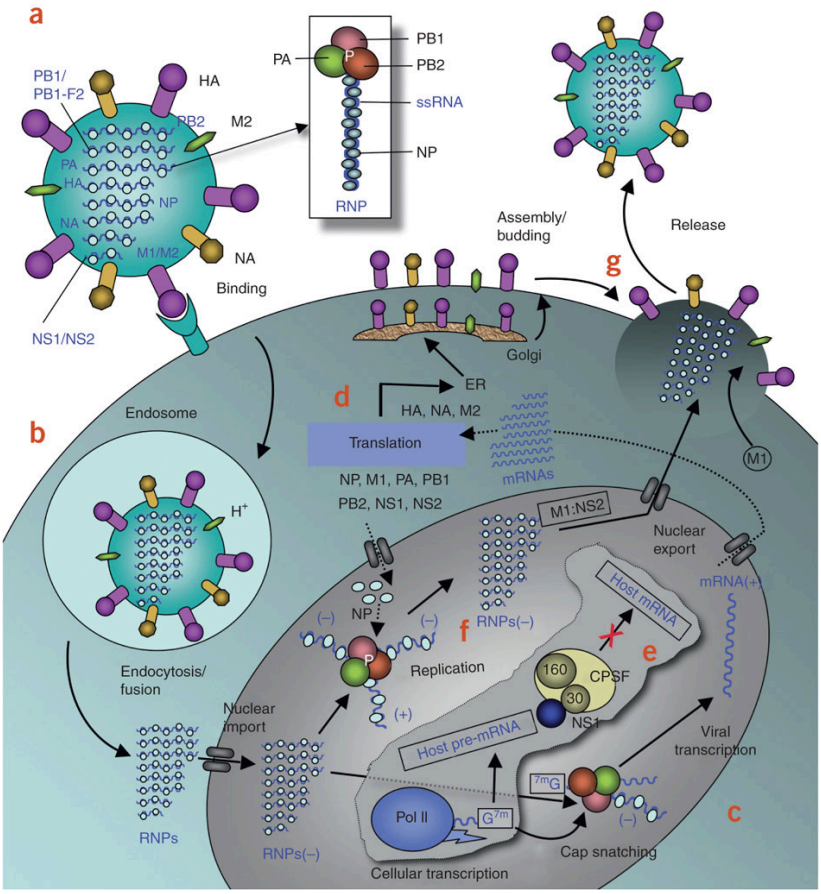


Fig. 5 Replication cycle of influenza A virus[10]

HA is a homotrimer that forms spikes on the viral lipid membrane (Fig 5a). To initiate infection and replication, the HA protein attaches to the host cell

receptors, containing terminal α-2,6 linked or α-2,3 linked sialic acid (Fig. 5b) [10]–[12].

The precursor of HA, HA0, is composed by two subunits, linked by a disulphide bond: HA1, which contains the receptor binding domain, and HA2, which contains the fusion peptide[13].

Two major linkages are found between sialic acids and the carbohydrates that are bound to glycoproteins: α(2,3) and α(2,6).

These glycans, terminated by α–D-N-acetyl neuraminic acid; Neu5Ac, act as receptors for the HA molecules, and the binding specificity of the viral surface HA to sialylated glycan receptors on the host cell surface is one of the many factors that govern adaptation of influenza to the human host. Viruses from avian and equines recognize the α (2,3) linkages, whereas viruses from human recognize specifically the α (2,6) linkage, even if humans have both type of receptors, α (2,6) in the upper respiratory tract and α (2,3) in the lower respiratory tract. Pigs, on their tracheal epithelial cells, have both (α (2,3) and α (2,6)) types of receptors [12]. This explains also the importance of swine being a good mixing vessel for avian and human influenza viruses, as a consequence producing dangerous pathogenic viruses. Moreover, since sialic acid-presenting carbohydrates are present on several cells of the organism, the binding capacity of the HA explains why multiple cell types in an organism may be infected[7]. After attachment, the virus enters through the receptor-mediated endocytosis of the virus particle. The endocytosis formed vescicles with viral particles fuse to endosomes in the cytoplasm, and during this step a low pH (5) is needed. The low pH in the endosome triggers a conformational change in the HA protein that leads to fusion of the viral and endosomal membranes. The low

pH also triggers the flow of protons into the virus via the M2 ion channel, thereby dissociating the vRNPs (viral ribonucleoproteins) from M1 matrix proteins. The vRNPs that are released into the cytoplasm are transported into the nucleus [14] where viral RNA synthesis takes place (Fig. 5c).

Viral RNAs (vRNAs) are used as templates by the viral RNA polymerase, and produce two kinds of positive-strand RNAs [15]. Complementary RNAs (cRNAs) are full-length copies of the vRNAs, and mRNAs are capped and polyadenylated [16]. The addition of the poly (A) tail occurs at a stretch of uridine residues close to the 5' end of the vRNAs. Amplification of the vRNA is done via copying of the full-length cRNAs into new full-length vRNA molecules [17]–[21](Fig. 5d-5e). Therefore, the viral mRNAs are transported from the nucleus into the cytoplasm, and it is than translated into proteins (Fig. 5f). The translated viral proteins that are needed for replication and transcription are imported back into the nucleus again. After that, newly synthesized RNPs are exported to the cytoplasm for packaging. Viral HA, NA and M2 are matured in the Golgi apparatus and collected at the plasma membrane, where with the help of M1, the production of viral particles begins. As a consequence, budding starts and the progeny virus is released from the cell by the activity of NA that destroys the receptors of the host and viral membranes. This is necessary for progeny virions to be released from the cell surface (Fig.5g) [22][23].

1.2 Influenza A virus

Highly pathogenic avian influenza, or, as it was originally called, “fowl plague”, was initially recognised as an infectious disease of birds in chicken in Italy, in 1878; but only after 1955, Schäfer characterised these agents as influenza A

viruses[7]. In the natural reservoir hosts of avian influenza viruses, wild water birds, the infection generally runs an entirely asymptomatic course as influenza A virus biotypes of low pathogenicity coexist in almost perfect balance with these hosts[24][25].

When low pathogenic avian influenza virus (LPAIV) strains are transmitted from avian reservoir hosts to higly susceptible poultry species such as chickens and turkeys (a trans-species transmission step), only mild symptoms are induced in general. Moreover, in some cases the poultry species are able to support several infection cycles, and these strains may undergo a series of mutation events resulting in adaptation to their new hosts. Influenza A viruses of the subtypes H5 and H7 not only run through a host adaptation phase but may have the capability to saltatorily switch by insertional mutations into a highly pathogenic form (highly pathogenic avian influenza viruses, HPAIV) inducing rapidly fatal disease.

HPAI in poultry is characterized by a severe illness of a short duration and a mortality approaching virtually 100% in vulnerable species. Before 1997, HPAI was fortunally a rare disease, with only 24-recorded outbreacks since the 1950s.

Recently, avian influenza acquired world-wide attention, when a highly pathogenic strain of the subtypes H5N1, which probabily emerged before 1997 in Southern China, become a regularly disease wich affect poultry throughout South East Asia and unexpectedly 'transversed interclass barriers'[26] when transmitted from birds to mammals.

During the 20th century, three different influenza pandemics occurred and their mortality impact ranged from devasting to moderate and mild.

The first influenza pandemic of the 20th century spread more or less simoultaneusly in different waves during 12 months between 1918 and 1919 across Europe, Asia and North America[27]. It was the worst pandemic in history, which killed more people that World War I, and at least 50 million people died [28]. The 1918 pandemic was caused by a H1N1 virus of apparent avian origin [29], whereas the subsequent pandemic strains- H2N2 in 1957 and H3N2 in 1968 were reassortant viruses containing genes from avian viruses.

During the spring of 1918, the first wave started, it was highly contagious but not particularly deadly; only the second wave, which began in September 1918, spread the deadly form of pandemic.

The 1918 virus (H1N1) was extremely virulent and caused many deaths through secondary bacterial pneumonia. Symptoms in 1918 were so unusual that, initially, it was misdiagnosed as dengue fever, cholera or typhoid. The majority of the deaths in the 1918 pandemic occurred in the young people between the second and fourth decates of life [30]. The virulence and lethality of the 1918 virus were theorized to be due to several factors: the preferred binding of the virus in the human respiratory epithelial cells, the adaptation that may have occurred in an intermediate host such as a pig, the enhanced cytokine and chemokine activation, and the lack of humoral immunity in the human host.

The complete genotype of the H1N1 influenza virus, responsible for the 1918 influenza pandemic (Spanish influenza), has been recreated by genetic enginering using a sample obtained from World War I soldier and an Alaskan native woman preserved under the frost in the Alaska area, in this way viral samples were used for genetic sequencing [27][30]–[32].

In 1957 began the second pandemic, which was caused by H2N2, a clinically milder virus than the one responsible for the 1918 pandemic. In addition, the world was much better prepared to cope to the pandemic[33]. Mortality showed a more characteristic pattern, similar to that seen in seasonal epidemics, with most excess deaths confined to infants and the eldery. The global excess mortality of the 1957 pandemic has been estimated at 1-2 million deaths.

The third pandemic started in 1968; it was as well a mild pandemic, but brought its own set of special epidemiological surprises. The mortality impact was not even particularly severe compared to the severe H3N2 epidemic in 1967-1968, as well as two severe H3N2 epidemics in 1975-1976 and in 1980-1981. The death toll has been estimated to be around 1 million, and in the U.S., nearly 50 % of all influenza related deaths occurred in the younger population under 65 years of age. Sero-archaeological studies showed that most individuals aged 77 years or older had H3 antibodies before they were exposed to the new pandemic virus and that the pre-existing anti-H3 antibodies might have protected the elderly during the 1968 H3N2 pandemic[7][33].

1.3 A brief overview of the current situation

Major pandemics have occurred throughout history at an average of every 30 years and there is a general consensus that there will be another influenza pandemic[7]. Infectious diseases are the result of a conflict of interest between macroorganisms and microorganism. To become a pandemic strain, an influenza virus must comply with a series of requirements. It has to: (1) enter

the human body and replicate there, (2) cause illness in humans, and (3) be easily transmittable between humans.

In theory, it has to be more pathogenic than other competing influenza strains. A good adaptation is the prerequisite for success: adaptation to human cells, the capability to produce new offspring, in addition to making the individual cough and sneeze to spread the offspring viruses. The clue to success is virulence and novelty: if the virus is a new virus, most humans will have little or no protection at all.

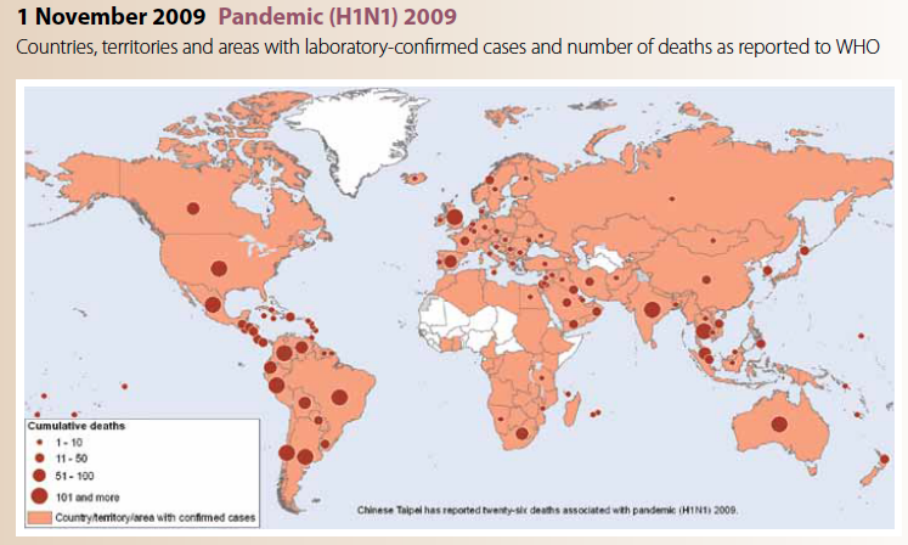


Fig. 6. Pandemic H1N1 2009. Countries with laboratory confirmed cases.

In 2009, a strain of influenza A(H1N1) virus, emerged, spread across the world and caused the 2009 H1N1 pandemic. In November 2008, more than 206 countries in the world have reported laboratory-confirmed cases of pandemic influenza H1N1 2009, including 6250 deaths[34]. This pandemic A(H1N1)2009 virus has been idealy circulating across the globe since 2009, and is now

established in human population as seasonal influenza virus. In February 2013, a novel reassortant influenza A (H7N9) birds was identified in eastern China, which by 30 April had spread to more than 11 provinces and municipalities (Fig. 6)[33][35]. As of 28 January 2014, the case fatality rate of all confirmed cases is 22%, but many cases are still hospitalized. Of all cases, 67% were male.

The median age of reported cases is 58 years old and that the avarage of fatal cases is 66 years old[36].

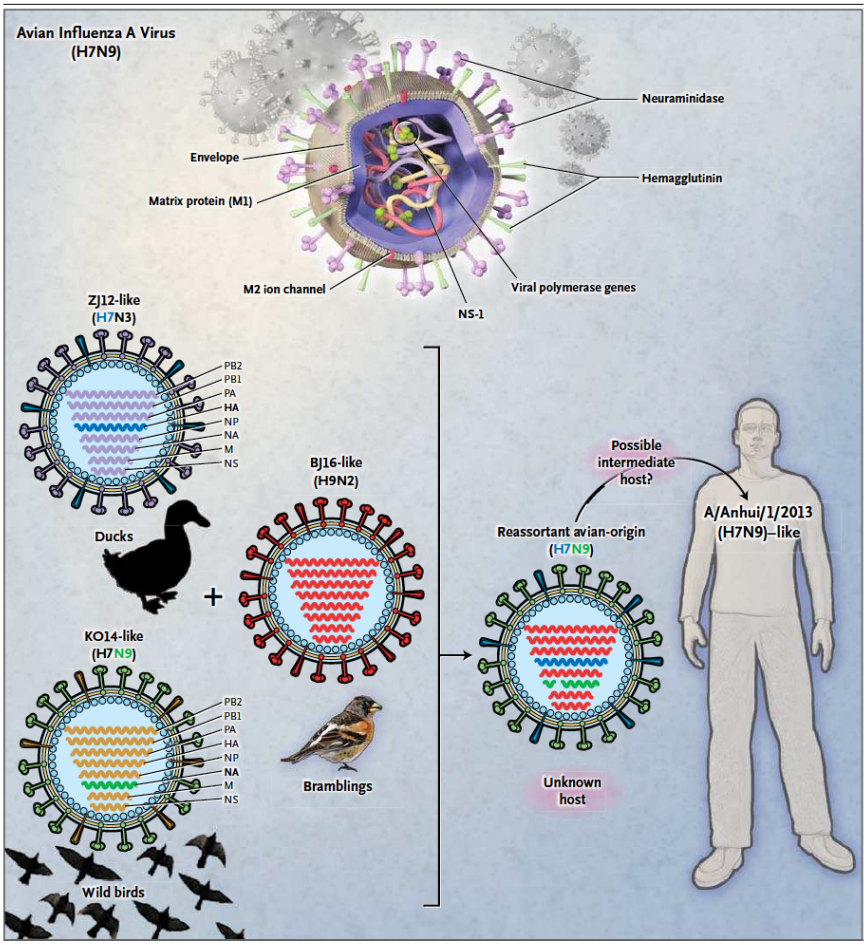


Fig. 7 Genetic evolution of the H7N9 new virus strain

Avian influenza is a subtype of influenza viruses that have been detected in birds in the past and this particular A(H7N9) virus had not previously been seen in either animals or people until 2013 in China. The HA gene is most similar to that of A(H7N3) viruses detected in ducks in Eastern China.

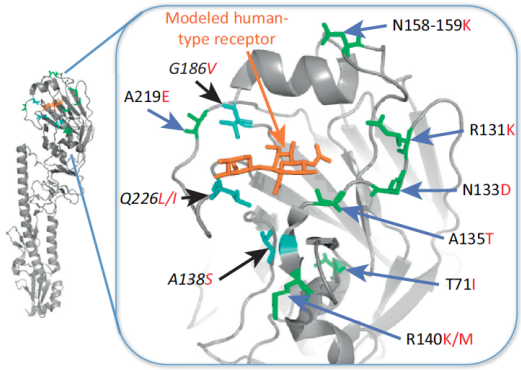


Fig. 8 Aminoacid changes in the HA of viruses recovered from contact ferrets in the human-infecting H7N9 virus groups. The 3d structure of A/Anhui/1/2013 (H7N9) HA is shown in complex with human receptor analogues. Mutations shown in cyan are know to increase the binding of avian H5 and H7 viruses to human-type receptors. Mutations that are emerged in HA of human-infectin H7N9 viruses during replication and/or transmission in ferrets are shown in green[37].

The NA gene is most similar to N9 NA genes from viruses circulating recently in domestic ducks in China and Korea. The six internal genes are derived from influenza A(H9N2) viruses circulating in poultry in eastern Asia. Sequence analyses have shown that the genes of the H7N9 viruses from China are of avian origin, but with signs of adaptation to mammalian species (Fig. 7).

In fig. 8 are shown the mutations that are known to increase the binding of avian H5 and H7 viruses to human-type receptors (A138S, G186V and Q226L/I). Others studies[37] using sequence analyses of viruses obtained from infected and contact animals identified other mutations in HA (R131K, T71I and A135T)

that seems to increase the ability of the virus to bind human receptor. Positions 131 and 135 are located near the receptor-binding pocket.

Recent works[38]–[43] have highlighted that these viruses exhibit high replicative ability and limited transmissibility in mammals, but have acquired mammalian-adaptating amino acid changes, and they may reassort with circulating human viruses. Although no efficient human-to-human transmission has occurred, the emerging H7N9 avian influenza virus poses in any case a potentially high risk to humans, for more than one reason: first of all, this new virus is able to bind both avian-type (α2,3-linked sialic acid) and human-type(α2,6-linked sialic acid) receptors[44], it has a high growth ability and furthermore human population does not have the pre-existing immunity to the H7N9 virus.

1.4 The receptor binding site

As already discussed, the cellular receptors for influenza virus are the glycoproteins and glycolipids having glycans terminated at their non-reducing end by a sialic acid (Neu5Ac). They bind in a shallow depression at the top of the HA molecule, which is composed of residues conserved in all subtypes of influenza. Several studies have been done using X-ray of complexes with other receptor analogs, confirming the model in which the structure and orientation of sialic acid are essentially identical irrespective of the receptor analog of these complexes. One side of the pyranose ring faces the base of the site, and the axial carboxylate, the acetamido nitrogen, and the 8 and 9- hydroxyl groups face into the site and form hydrogen bonds with conserved side chain or main-chain polar atoms.

Specifically, a conserved serine in position 136 forms a hydrogen bond with the carboxylate, which is also hydrogen bonded to the amide of peptide bond at position137, histidine 183 and glutamic acid 190 form hydrogen bonds with the 8-hydroxyl group. The 5-acetamido nitrogen forms a hydrogen bond with the carbonyl of peptide bond 135, and the methyl group of this substituent is in van der Waals contact with the six-membered ring of tryptophan 153.

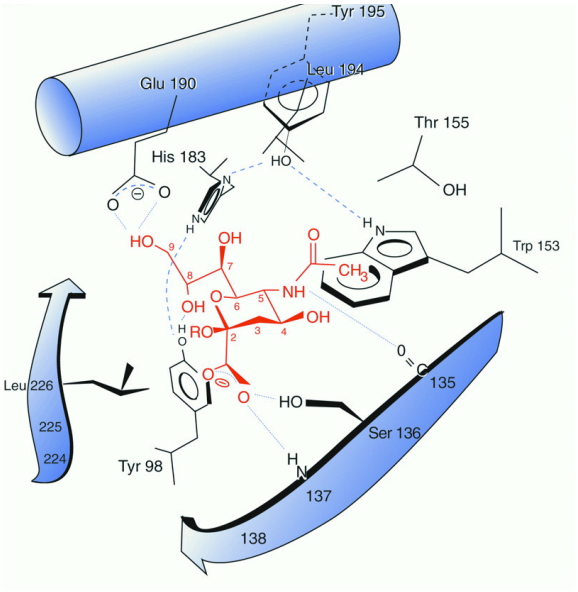


Fig. 9 Hemagglutinin receptor binding. Schematic diagram of the receptor-binding site of an HA-receptor analog complex showing the hydrogen bonds that in wt HA form with bound sialic acid.

The 7-hydroxyl group and acetamido carbonyl hydrogen bond to each other and form van der Waals contacts with leucine 194[12].

The most commonly glycans used to represent avian and human receptors are LS-tetrasaccharide a (LSTa; Neu5Acα2–3Galβ1-3GlcNAcβ1-3Galβ1-4Glc) and LS-tetrasaccharide c (LSTc; Neu5Acα2–6Galβ1-4GlcNAcβ1-3Galβ1-4Glc), and the

X-ray co-crystal structures are often solved using these oligosaccharides. The conformation of LSTa and LSTc have been characterized primarily by the glycosidic torsion angles of the terminal sialic acid linkage[45].

In the case of the Neu5Ac α(2,3)-Gal linkage in LSTa complexed with avian-adapted HA, the torsional angle φ (C1-C2-O-C3) is $\sim$ 180° and is described as the *trans* conformation. While the Neu5Ac-α(2,6)-Gal linkage in LSTc complexed with human-adapted HA, φ (C1-C2-O-C6), is $\sim$ -60°, or a *cis* conformation. This *cis* and *trans* definition of glycan receptor conformation enabled distinguishing key contacts of residues within the RBS to either LSTa or LSTc. Even if the *cis* and *trans* definition of glycan conformation has been useful to characterize the distinct interactions with the terminal Neu5Ac α(2,3)-Gal or Neu5Ac α(2,6)-Gal motif, this definition does not fully describe HA binding to a range of structurally diverse glycans present in human respiratory cells and tissues[46].

In order to extend the conformational analysis beyond the terminal sialic acid linkage to describe overall topology and dynamics of the glycan receptor upon binding to the RBS of avian and human-adapted HAs, [47][48] a parameter θ has been defined to measure the angle between the Neu5Ac, the penultimate Gal, and the third GlcNAc sugar (Fig. 10 B). Using the θ parameter it is possible to classify the ensemble of conformations sampled by the avian and human receptors within the HA binding site. In the case of avian receptors, the conformations sampled by the Neu5Ac α(2,3)-Gal linkage (keeping the sugar anchored) and the sugars beyond this linkage (at the reducing end) span a region on the binding surface of HA that resembles a cone, characterized by a θ angle >100°. The different conformations sampled by Neu5Ac α(2,6)-Gal

linkage and the sugars beyond this linkage (at the reducing end) span a wider area on the HA binding surface.

In this case there are two portions of the RBS receptor contacts that can be described in one case as a cone-line surface, whereas the other portion is more correctly described as umbrella-like and is characterized by θ angle<100°. It was shown that the umbrella-like topology is predominatly adopted by human receptors that posses at least four sugars including Neu5Ac, whereas the cone-like topology was shown to be adopted by avian and human receptors[46].

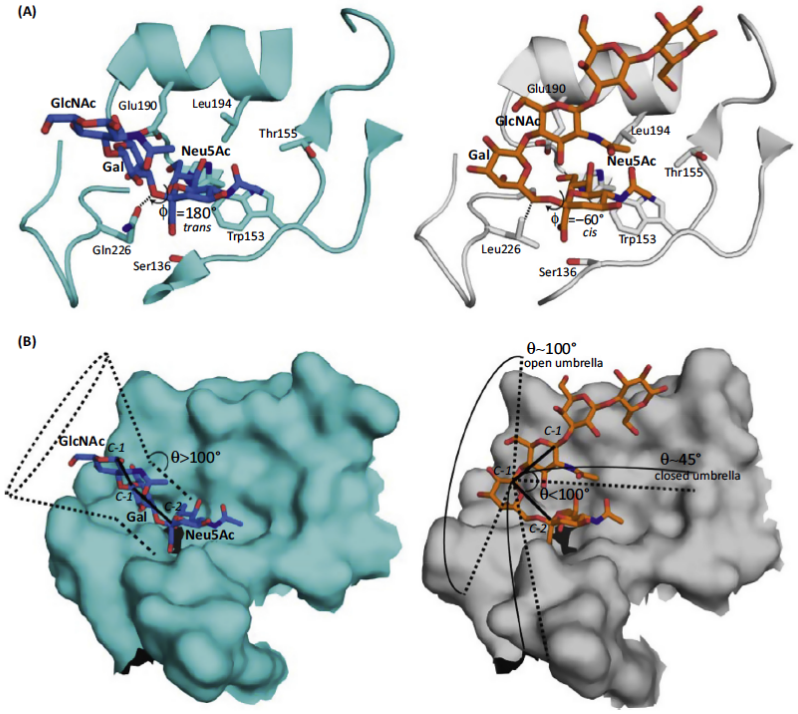


Fig. 10 Glycan receptor conformation and topology in hemagglutinin (HA) receptor binding site (RBS). (A) Left panel shows the trans conformation adopted by LSTa in the RBS of avian-adapted H3 HA (PDB ID:1MQM). Right panel shows the cis conformation adopted by LSTc in the RBS of pandemic H3 HA (PDB ID:2YPG). (B) Left panel shows the complex avian HA-LSTa glycan complex where θ is the angle between C-2 atom of Neu5Ac, C-1 atom of Gal, and C-1 atom of GlcNAc. Right panel shows the pandemic H3 HA-LSTc glycan complex where θ<100°.

1.5 Receptor binding specificity and its biological significance

As previously described, two major linkages between sialic acid (Neu5Ac) and the penultimate galactose (Gal1) residues of carbohydrate chains are found in nature: Neu5Ac α(2,3)-Gal and Neu5Ac α(2,6)-Gal. Different HAs have different recognition specificities for these linkages, and for example A/Hong Kong/68 , X-31, HA binds preferentially to sialic acid in the α(2,6) linkage. In contrast, the HA of a mutant of X-31, with a single mutation L226Q HA, has a greater affinity than the wild type HA for α(2,3)-linked sialic acid. The single aminoacid sequence difference, L226Q, between wild type and mutant HAs defined the location of the receptor binding site.

Two related aspects of linkage recognition specificity have been extensively researched for their biologic significance. The first associates to differences observed in the recognition specificities of HAs from viruses infecting different species[49]–[52]. These differences may be important in limiting the transfer of viruses between species and were analysed especially in relation to the emergence of new pandemic viruses in humans. Experiments with linkage-specific lectins have shown an abundance of sialic acid in α(2,6) linkage in human lungs and of sialic acid in α(2,3) linkage in bird intestine; in pig, both linkages were detected in respiratory tract cells[50]. Moreover, mucins from human lung are reported to be rich in α(2,3)-linked sialic acid and in contrast, equine fluid may be rich in non-specific inhibitors containing sialic acid in α(2,6) linkage[53].

The molecular basis of the differences in receptor-binding specificity between H3 subtype viruses isolated from humans and avians has been addressed by sequence analysis[54]. Residue HA 226 again appears to be involved, and HA

228 has also been linked with specificity differences. The HAs of viruses from humans that preferentially recognize the α(2,6) linkage contain leucine and serine at 226 and 228; whereas the avians HAs contain glutamine and glycine at the same positions[55]. The subtype H1 differs in this; avian and human isolated viruses also recognize α(2,6) and α(2,3) lingakes, respectively, but both of their HAs contain glutamine at position 226 in HA1 and glycine in position 228 HA1. Other residues in HAs of this subtype, including 186 (Proline in avian and Serine in human) and 225 (Glycine in avian and Aspartate in human), influence receptor specificity. Inside the H3 subtype, there are others residues, in addition to 226 and 228, which have been shown to affect linkage recognition specificity[12].

The receptor binding specificity of human and avian influenza viruses suggest that avian viruses need to acquire the ability to recognize human-type receptors to cause a pandemic[55], but it is clear that the influenza A viruses remain as “unpredictable” pathogens. The concept of switch in glycan receptor specificity needs to be defined and interpreted carefully to enable it to be a useful tool for surveillance and strain characterization; this because amino acid mutations that confer specific receptor binding properties to a particular strain and subtype of HA do not always confer the same properties to a HA of a different virus. Until now almost only X-ray crystallography has been used in order to characterized the binding between HA and its human and avian receptors[56]–[62], even if due to the conformation complexity of the HA-glycan complexes, the structures were often resolved with a trisaccharides instead a pentasaccharide. Only recently NMR (Nuclear Magnetic Resonance) and Molecular Dynamic (MD) are emerging as complementary techniques able

to detect the dynamics of glycan receptor motion as it transition from unbound to bound states with HA.

2. A brief introduction to Nuclear Magnetic Resonance

Nuclear Magnetic Resonance, or as usally called by its users NMR, was discovered in 1945 by Bloch, Hansen and Packard at Standford University and by Purcell, Torrey and Pount at Harvard University and for their discoveries Purcell and Bloch shared the Nobel Prize in Physics in 1952.

Even if discovered by phisichian, NMR spectroscopy is used principally by chemist and biochemists to investigate the properties of organic molecules, but it is also a powerful tool for studying the structure, function and dynamics of biological macromolecules[63][64].

2.1 Overview of the concepts

An intrinsic property of many nuclei is the spin angular momentum, which give rise to magnetic moments[65]. Electrons, ^{1}H, ^{13}C and ^{15}N nuclei are characterized by a common spin quantum number s=1/2 and magnetic spin quantum number I_s= ±1/2. Most biological NMR experiments utilize one or more of the above nuclei for investigation.

There are two possible states for an s=1/2 nuclei: spin up (denoted ↑ or α) and spin down (denoted ↓ or β) associated I_S= +1/2 and -1/2, respectively (Fig.11). When no magnetic field is present, states with different values of magnetic quantum number I_S are degenerate and their energy is equal. However, when an external applied field is present the different states are no longer degenerate. As a result, the energy levels are split in two when a field is present [66].

The difference in energy is given by: $\Delta E = E\alpha - E\beta = h\nu = -\gamma\hbar B_0$

where $E\alpha$ and $E\beta$ denote the energy for the α and β states, γ is the gyromagnetic ratio (the proportionality constant between ΔE and $B_0 I_S$ [66]), $\hbar$ is the Planck constant h divided by 2π, B_0 is the steady magnetic field applied and I_S is the quantum magnetic number.

The application of a magnetic field B_0 to a particle with a magnetic moment $\gamma\hbar I_S$ generate an energy splitting $\Delta E = -\gamma\hbar B_0$ where the corresponding frequency $\Delta E = h\nu_L$ is usually called Larmor Frequency. In a classic description of the NMR event the Larmor Frequency corresponds to the processing rate of the spin angular momentum around the direction of the know component (I_S), defined by the direction of the external magnetic field. The rate at which the vectors precess depends on the nucleus type and is proportional to the applied magnetic field[66]. The Larmor frequency, ν_L, takes the value:

$$\nu_L = -\frac{\gamma B_0}{2\pi}.$$

At equilibrium the population distribution between the two states (α or β) follows the Boltzmanns distribution:

$$\frac{N_\beta}{N_\alpha} = e^{-\Delta E/k_b T}$$

where N_α and N_β are the number of spin in each state, k_b=Boltzmanns constant and T is the temperature (Fig. 11).

At equilibrium, for nuclei with positive γ, N_α is slightly more populated than N_β, and vice versa for nuclei with negative γ. This excess of nuclear spins gives rise to a bulk magnetisation Mz along the +z-axis (defined axis for equilibrium net magnetisation). In the xy-plane, however, all the vectors are randomly distributed and their individual contributions cancel out[65]–[67].

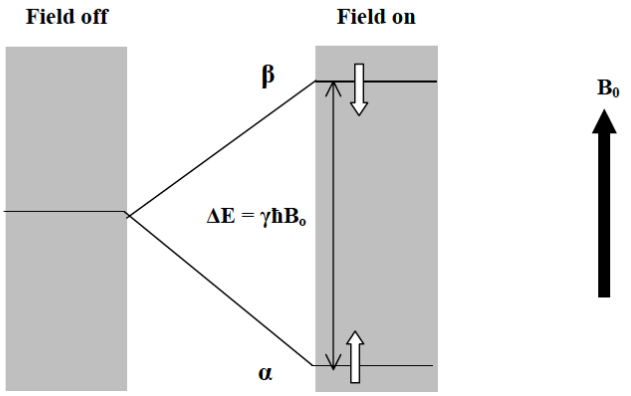


Fig. 11 The nuclear spin energy levels of spin -1/2 nucleus in a magnetic field.

Nuclear magnetic resonance occurs when the nucleus changes its spin state driven by the absorption of a quantum of energy: by applying radio frequency (RF)-pulses (Larmor frequency) to a system, transitions between energy states can be induced. Traditionally, the effect of applying the pulse for a defined duration (correspond to the $\pi/2$ pulse), is pictured as the net magnetisation vector along the z-axis flipping down onto the xy-plane. In addition, due to phase coherence, net magnetisation in the xy-plane can now be observed[68]. Phase coherence is the alignment of all magnetisation vectors in the xy-plane: their contributions no longer cancel out. The evolution of magnetisation (which is a product of decay and resonance frequency) is measured.

There are essentially two relaxation pathways in which the system returns to the original state that is measured in an NMR experiment: 1) Relaxation from the xy-plane to the z-axis and 2) relaxation within the xy-plane.

The first is called spin-lattice relaxation, longitudinal relaxation times or T_1 relaxation and occurs because there is an exchange of energy between the spins states and the surrounding medium. It is also referred to as spin-lattice

relaxation time, because the excess energy is released to the "lattice" (the surrounding system of energy states) when spins revert from the β to the α state.

The second relaxation pathway is called T_2 relaxation or transverse relaxation, and occurs with exchange of energy between different nuclear spins. This is also called spin-spin relaxation time to reflect that it involves the relative orientations of the spins. It is useful to think of T_1 processes as affecting the lifetimes of population of spin energy levels, while T_2 processes affect the relative energies of the spin levels rather than lifetimes.

Through the two relaxation pathways, the system relaxes back to equilibrium at a rate dependent on the time constants T_1 and T_2. The signal produced is called FID (Free Induction Decay), which is the superimposition of the resonance frequencies of all spins in the molecule as a function of time, F(t).

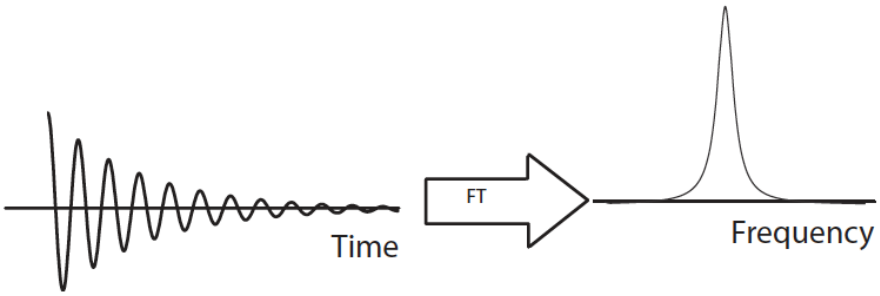


Fig. 12. Schematic description of how the shift from time-domain to frequency-domain by Fourier Transform is achieved for a NMR spectrum.

By applying the Fourier transform, a mathematical operation which allows to translates a function in the time domain into the frequency domain, it is possible to obtain the corresponding spectrum F(ω), as intensity as a function of frequency (Fig. 12) [65]–[67], [69], [70].

J-coupling

The *J*-coupling, or indirect dipole-dipole interaction, is an interaction between two nuclear spins mediated by chemical bond electrons. Two spins have a measurable *J*-coupling if they are connected through a small number of chemical bonds, including hydrogen bonds[70]. The interaction is measured by the scalar coupling constant $^{n}J_{ab}$, where n denotes the number of covalent bonds separating the two nuclei a and b. With scalar coupling (J-coupling), magnetisation is transferred through bonds: all the spins influence each other in a predictable way, and this can aid in determination and labelling of functional groups. *J*-coupling, differently from the chemical shift, is independent of the external magnetic field. Couplings between nuclei cause splitting of the NMR-signal. J-values for n=1-3 are dependant not only on internuclear distance but also on the number of bonds between the two nuclei and or the geometry of the molecule. As such, one can obtain possible parameters for structure calculations from scalar couplings[68]. ^{3}J-couplings (vicinal couplings) are particularly useful in this respect. The value is dependent on substituents, the distance between the two carbon atoms involved, H-C-C-H bond angle, along with torsional and/or dihedral angles, with the latter angle being of greatest interest [68].

^{3}J-couplings vary in magnitude in a bond-angle dependent manner. The relationships between the coupling constant J and the dihedral angles ϕ are given by the Karplus equation [70][71]. Consequently, J-couplings can be used to measure dihedral angles and thus conformational arrangements around bonds.

2.2 NMR spectroscopy and carbohydrate analysis

Aim of this paragraph is to briefly elucidate some NMR techniques and experiments useful for carbohydrate characterization and used in this thesis for the characterization of the pentasaccharides.

Carbohydrates are a group of compounds with vast structural and chemical diversity but with limited chemical shift dispersion in NMR spectra, which makes their study by NMR challenging and interesting. The assignment of ^{1}H and ^{13}C resonances to their corresponding protons and carbon atoms are usually achieved using a combination of 1D and 2D NMR experiments. ^{1}H-1D experiments are usually the starting point. In spite of most carbohydrate resonances are concentrated in the region between 3 and 4 ppm, anomeric proton signals (4.3-5.8 ppm) are usually well-resolved and can be used to estimate the number and molar ratio of different monosaccharides present in the sample. In order to resolve the strong signal overlapping, the uses of bidimensional (2D) experiments allow to better resolve the spin systems and to identify the nature of the corresponding monosaccharide. The *COSY* (Homonuclear **CO**rrelated **S**pectroscop**Y**) experiment provides NMR spectra in which each proton correlates with another one through a scalar coupling. The signals on the diagonal correspond to those of a normal one-dimensional spectrum, while the cross peaks show the correlations between coupled protons. *TOCSY* (**TO**tal **C**orrelation **S**pectroscop**Y**) experiment, although similar to COSY, differentiates by the fact that, each nucleos is correlated not only with directly coupled atoms but also to the rest of same spin system, regardless of whether they are themselves coupled to one another. In case of

oligosaccharides each residue is an isolated spin system that can be revealed and assigned to the type of monosaccharide[72]. In the *NOESY* (**N**uclear **O**verhauser **E**nhancement **S**pectroscop**Y**) experiment, differently from the previously discussed techniques in which the transfer of magnetization takes place between nuclei that have scalar coupling, magnetization is transferred through space by dipole-dipole interaction [73].

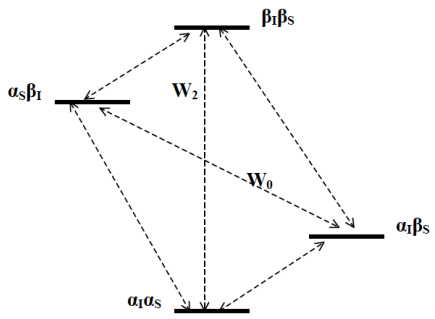


Fig. 13 Energy level diagram for a two-spin system, I and S, showing definitions of transition probabilities and spin states.

The rate of cross-relaxation (relaxation between two spins), and thus the NOE enhancement factor, is among others dependant on the distance between the coupled nuclei and the motion of the molecule. Cross relaxation leads to changes in the population of the two coupled spins. This change of population can again lead to peaks with stronger/smaller intensities as a result of the NOE-effect. Equalizing the population difference (saturation) of a spin S strongly affects the longitudinal magnetisation of a coupled spin I caused by the increased population difference of this spin. At this stage the system is not in a state of equilibrium and returns to equilibrium by relaxation mechanisms, mainly through dipole-dipole interactions. NOE and dipole-dipole relaxation mechanisms are thus closely connected. During cross relaxation, magnetism is

transferred through dipolar couplings between the interacting nuclei, and the intensity of NOE-peaks is proportional to r^{-6}, where r is the internuclear distance[65][66][70][74][75]. In case of carbohydrates, inter-residues noe's are often used to sequence the oligosaccharidic chains, as well as to characterize the glycosidic linkage geometries and residue conformations.

Finally, the *1D Sel-TOCSY and 1D Sel-NOESY* are the one-dimensional analogues of more conventional 2D experiments which allow greater digital resolution and thus a more detailed insight into the fine structure; these equate to high-resolution slices through the related 2D experiment at the shift of the selected spin. They may also benefit by being quicker to acquire and analyse. These methods may provide faster and more detailed answers when a specific structural question is being addressed[66]. These types of experiments, for their purpose, use selective pulses in order to excite or invert single resonances that can be either singlests or multiplets [76].

The second group of experiments, useful for the structure determination of carbohydrates, is called heteronuclear, and correlates two or more spins via single or multiple bond couplings. The *HSQC* (**H**eteronuclear **S**ingle **Q**uantum **C**oherence) experiment provides NMR spectra where the correlations between the proton atoms and carbons (or nitrogens in case of protein) directly linked to them can be observed. The larger chemical shift dispersion for carbon signals is an advantage in structural assignments. Since the ^{13}C anomeric signals appear in a characteristic region between 90 to 110 ppm, HSQC can provide a picture of the number and type of different monosaccharides in the sample. Multiplicity *edited-HSQC* experiments can be used to discriminate CH and CH_3 carbons from CH_2 carbons. In the *HSQC-TOCSY* experiment each crosspeak is correlated, in the proton dimension, to all atoms belonging to the

corresponding spin system. As a consequence, also each carbon of the same spin system is correlated to each other, facilitating the monosaccharide assignment of strongly overlapped systems.

The *HMBC* (**H**eteronuclear **M**ultiple-**B**ond **C**orrelation Spectroscopy) experiment correlates protons over ranges of 2-4 bonds. It is useful to provide assignments for non-protonated centres and to produce unambiguous carbon assigments when proton resonances exactly overlap[66], [77]–[79]. It is also largely used to find correlations across glycosidic linkages to determine the linkage between adjacend sugars. Finally the *J-HMBC* experiment is used to measure the magnitude of long-range ($^3J_{CH}$) couplings also in molecules at natural aboundance. Since ^{3}J-couplings vary in magnitude in a bond-angle dependent manner, their measurement, thanks to the relationships between the coupling constant value and dihedral angle given by the karplus equation, allows to determine the geometry of glycosidic linkages [70][71].

2.3 NMR as a tool for interaction studies

Molecular recognition lies at the heart of life processes. The interactions of receptors with natural products are key factors in many processes of biomedical importance, and therefore the understanding, at the molecular and atomic levels, of the diverse mechanisms by which these molecules are recognized by the binding site of receptors, antibodies and enzymes is of paramount importance[80]. From this perspective, knowledge of the structural, conformational and dynamic features of the molecular recognition processes between proteins and their ligands targets has the potentiality to rationalize

binding and selectivity features, and therefore to assist in the design of new molecular probes and of novel therapeutic agents.

NMR has become a powerful tool to monitor molecular interactions and to deduce features of recognition processes at different levels of complexity, from the perspective of the receptor and the ligand. Moreover NMR has the advantage to examine the sample in solution, in a condition quite similar to the physiological one, and the results obtained take into account the flexibility of the protein and the conformational changes that occur upon ligand binding.

2.3.1 NMR-Based protein-ligand interaction studies

The study of protein-protein and protein-ligand interactions in solution has recently become possible at an atomic level by new NMR spectroscopy experiments able to identify binding events either by looking at the resonance signals of the ligand or the protein. Ideally, a combination of these two informations allows obtaining a complete picture of ligand binding to a receptor. These methodologies can be applied to the lead generation in drug discovery[81], that is the identification of compounds that demonstrates specific activity against a therapeutic target, as well as to dynamic structure-activity relationships (SAR). Among the most important approaches, we can distinguish: (1) chemical-shift variations by titration experiments $^{1}H/^{15}N$-HSQC and $^{1}H/^{13}C$-HSQC which are used to identify the binding pocket of the receptor (2) Saturation Transfer Difference (STD) experiments, used to identify the binding motif of the ligand (3) transferred NOEs effect for detecting and characterizing bound conformation of the ligand [82],[83][73], (4) NOE pumping experiment used to suppress the signals of the unbound

molecules[84], (5) DOSY (Diffusion Ordered Spectroscopy) spectroscopy seeks to separate the NMR signals of different species according to their diffusion coefficient[85].

Characterization of fast/slow exchange by NMR

Using NMR it is possible to characterize several aspects of protein-ligand interactions: (1) determine the binding affinity, (2) location of ligand binding site on protein and (3) the structure of the protein-ligand complex. The knowledge of the strength of the interaction is foundamental for the study of protein ligand interactions. The affinity of ligands for protein targets ranges over several orders of magnitude, and it is represented by the dissociation constant K_d, where K_d is defined as:

$$K_d = [T][L]/[TL]$$

where T refers to the target protein, L the ligand and TL the complex concentrations.

In general, according to the different strenght of binding is possible to identify: (1) tight-binding interaction when K_d is less than 10^{-8}M, (2) moderate interaction for K_d approx between 10^{-6} -10^{-8} M and (3) weak interaction when the K_d is greater than 10^{-6} M.

The magnitude of K_d has an important effect on a number of NMR parameters and plays a significant role in determing what sort of information may be obtained for a particular protein ligand interaction.

If a small ligand binds to a protein with low affinity, and the exchange rate between the free and bound states is fast enough (K_{off} larger that the chemical shift difference) it is possible to apply the following equation:

$A=p_{bound}A_{bound}+p_{free}A_{free}$

in which its NMR osserable A equals the population-weighted average of A_{bound} and $A_{free.}$ [86].

The NMR monitors the different physical properties that exist between a ligand and a protein.

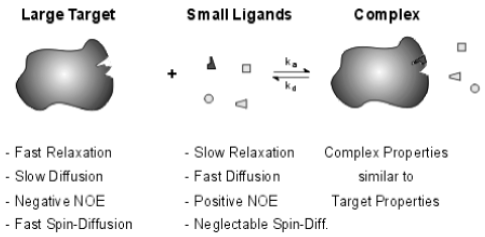


Fig. 14 Characteristic of the free ligand, target and the correspondent complex.

The exchange regime (slow, intermediate or fast) determines how a spectrum of a protein-ligand mixture changes (Fig. 16) during a titration and depends on K_d.

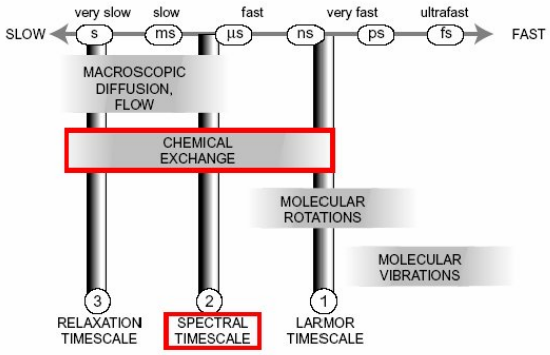


Fig. 15 Rapresentation of the chemical exchange and correspondent spectral timescale.

It is important to determine if the complex is in fast or slow exchange on the NMR timescale (Fig. 15), before starting to study the structure of a complex. If the complex is in a slow exchange, the signals of the free and bound state are

separated, even if this latter is usually very broad. If the complex is in fast exchange, only one signal is visible (Fig. 16) and its chemical shift is the weighted average of the free and bound state signals [86]–[88]. Finally, in the intermediate exchange the signals of the free and bound states appear broad and often not resolved.

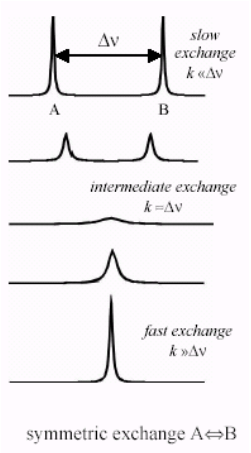


Fig. 16 Simulated spectra for two site exchange.

2.3.2 Saturation Transfer Difference (STD)

The saturation transfer difference (STD) method permits a fast detection of ligand binding to proteins, and allows an easy discrimination of ligands from non-binding molecules[80][82]. For this experiment, the chemical exchange between the bound and free states of the ligand should be fast on the relaxation time scale (K_d between 10^{-3} and 10^{-7}M). In STD experiment, assignment of the protein cross-peaks (only the ligand signals are monitored), and isotope labeling of the protein or the ligand are not required. STD

experiments require small quantities of receptor (<20 μM), and use large excees of ligand (ca. 100 molar equivalents versur the receptor).

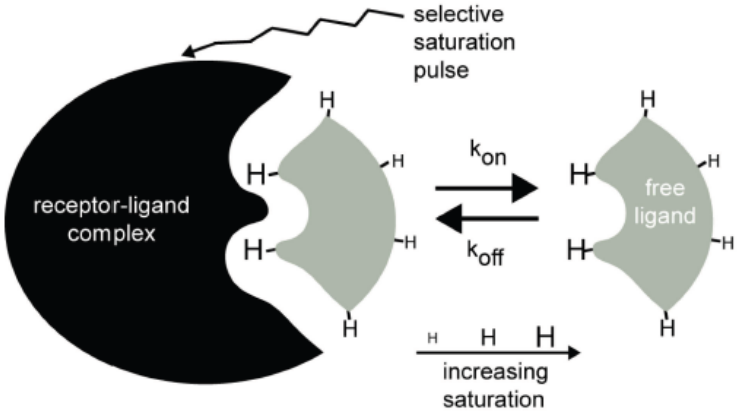


Fig. 17 Schematic view of the STD experiment. The protein protons are selectively saturated at a specific frequency, and the effect of the saturation is transmitted through the whole polypeptide. Any ligand interacting with the protein will also be affected by the saturation. During the irradiation time, the saturation is transferred to the bound ligand- first to the protons belonging to the ligand epitope, then to the rest of the ligand.

STD results from the difference of two experiments (Fig. 18).

In the first experiment, the "on-resonance" experiment, selective saturation of the signal of the receptor nuclei is achieved via a train of frequency-selective rf pulses. The rf train is applied to a frequency window that contains protein resonances, but not resonances from the ligand. The saturation propagates rapidly from the point of application to other protein nuclei via the vast network of intramolecular ^{1}H-^{1}H cross-relaxation pathways; this process of spin-diffusion is very efficient, due to the typically large molecular weight of the protein. Bound ligands pick up this saturation via inter-molecoular ^{1}H-^{1}H cross-relaxation at the ligand-protein interface. They then dissociate back into free

solution where the saturated state persists, due to the small free state R_1 values. At the same time, more “fresh” ligand exchanges on and off the protein while saturation energy continues to enter the system through the sustained application of rf. Thus, saturated free ligands accrue during the saturation time and this allows the “amplification” of the signal. For the on-resonance irradiation, frequency values around -1 ppm are used because no ligand nuclei resonances are found in this spectral region, whereas the significant line width of protein signals still allows selective saturation. If the ligand shows no resonance signals in the aromatic proton spectral region, the saturation frequency may also be placed there or even further downfield (δ = 11-12 ppm).

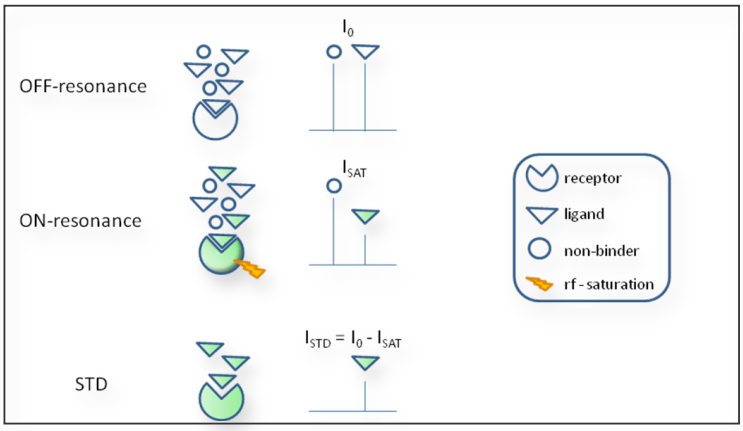


Fig.18 STD spectrum is generated by subctration of “on-resonance” spectrum from “off-resonance” spectrum.

In order to achieve the desiderated selectivity and to avoid side-band irradiation, shaped pulses are employed for the saturation of the protein signals.

Therefore, in the second experiment the irradiation frequency is set to a value that is far from any ligand and protein signal (off-resonance). The spectrum is

recorded and yields a normal NMR spectrum of the mixture. Subtraction of the on-resonance spectrum from the off-resonance spectrum leads to a difference spectrum, in which only proton signals that were attenuated by saturation transfer are visible. The intensity of STD signals is related to their distance with the protein surface, being higher for those protons that are in closer contact with the receptor.

Thus, it appears evident that the STD NMR is an excellent technique to show ligand-protein binding, as well as to characterize the binding epitope of the ligand[80][82][89].

2.3.3 Transferred NOESY (trNOESY)

NOE effects (NOEs) are extremely useful in determining the 3D structure of molecules in solution. The method has brought many benefits especially with regard to the structural analysis of proteins. When ligand molecules bind to receptor proteins, the NOEs undergo drastic changes, leading to the observation of transferred NOEs (trNOEs). These changes are the basis for a variety of experimental schemes that are designed to detect and characterize binding activity.

The observation of trNOEs relies on different tumbling times τc of free and bound molecules. Low or medium-molecular-weight molecules (Mw <1000-2000) have a short correlation time τc and, as a consequence, such molecules exhibit either positive NOEs, no NOEs, or very small negative NOEs depending on their molecular weight, shape, and the field strength. Large molecules, however, show strongly negative NOEs (Fig.19). When a small molecule (ligand)

is bound to a large-molecular-weight protein (the protein receptor molecule) it behaves as a part of the large molecule and adopts the corresponding NOE behaviour.

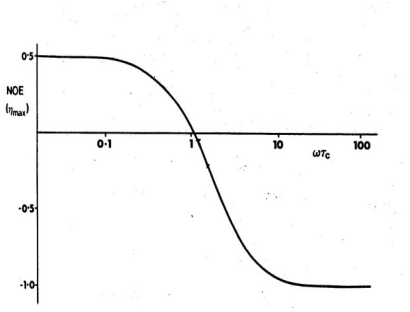


Fig. 19 . Dependence of maximum homonuclear NOE enhancement on $\boldsymbol{\omega\tau_c}$.

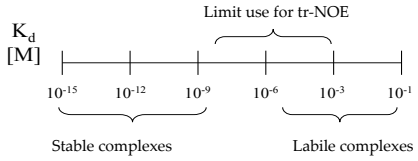


Fig. 20. Range of K_d used in tr-NOESY experiment

Ligand binding of a receptor protein can thus easily be distinguished by looking at the sign and size of the observed NOEs.

The estimated range of binding affinities that can be probed by trNOESY is $100nM \leq K_D \leq 1mM$[90](Fig.20).

Cross-peaks among protons cross-relaxing across the glycosidic linkages and within the monosaccharide units can be monitored in absence and in presence

of the protein, identifying variation of the conformation of the ligand in the bound state.

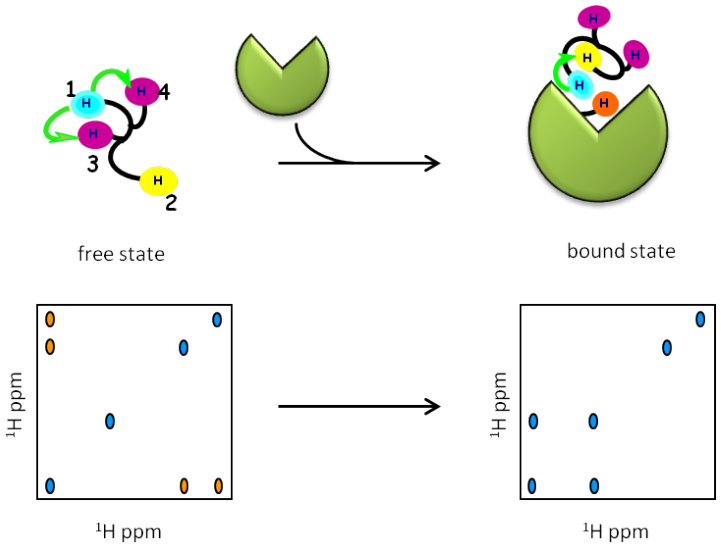


Fig. 21 Left. Schematic representation of a NOESY spectrum for a free ligand. Cross peaks and diagonal peaks have different sign. Right. Schematic representation of a trNOESY spectrum recorded for an exchanging ligand–protein system. Cross peaks and diagonal peaks have the same sign, as expected for a large molecule, thus indicating binding to the protein.

One of the major drawbacks of this kind of experiment is the possible presence of the spin diffusion effect, that is typical for high-molecular weight proteins[82]. In this case, the magnetization transfer can be mediated by other spins, including receptor nucleus, and by protons closed in space.

Thus, the trNOE method allows fast screening of putative binders in respect to a specific target and, at the same time, permits the knowledge of the recognized conformation of the ligand bound to the receptor, with considerable implications for a rational structure-based drug design.

3. A simplified introduction to Molecular Dynamics Simulation

The NMR spectra of a molecule or macromolecule in solution allow the direct observation of the magnetic properties (chemical shifts measurements, coupling constants, magnetization relaxation, ect.) that are related to the structural, conformational, and dynamic features of the molecule, and also to its interaction with neighbour species in the studied sample. In this context the adjective “structural” for a molecule is related to proprieties strictly dependent to the atom positions, while “conformational” aspects are interesting anytime the molecule exist in equilibrium between several conformers that can reversibly interchange without breaking bonds, by rotation around some of their bond axis. The dynamic behaviour of the molecule is the whole motion, usually decomposed in the internal, involving only internal degrees of freedom, and tumbling motion. The modeling approach allows a “bottom-up” description based on the general idea to reproduce by “numerical experiment” (computer simulation) the chemical and physical behaviour of our system being a molecule in solution or a macromolecule or a supra-molecular complex.

Computer simulations act as a bridge between microscopic length and time scales toward the macroscopic world of the laboratory: roughly speaking this approach provides a guess of the conformational and dynamic features of the macromolecule, the forces acting between atoms that belong to the same or to different molecules, and compares this model behaviour with the NMR experiments results. In particular the molecular mechanic and molecular dynamic approach are suitable to this scope.

Molecular dynamics simulation consists of the numerical, step-by-step, solution of the classical equations of motion, which for a simple atomic system belonging to a single or several molecules may be written in accord to the Newton second law of the classical mechanic

$$m_i \ddot{r} = f_i \qquad f_i = -\frac{\vartheta}{\vartheta r_i} U \qquad (1)$$

For this purpose we need to be able to calculate the forces f_i acting on each atom of mass m_i that is subject to an acceleration $\ddot{r}_i$. The forces are derived from an empirical potential energy $U(r^N)$ that normally is parameterised for a particular set of molecule or macromolecule or material. The parameters of this potential energy correspond to molecular structural parameter for a particular class of compound, such as bond equilibrium distances, angles, and torsional angles. There are also other parameters that quantify the ability of these degree of freedom to fluctuate around their equilibrium values, and represent the system energy cost to a deform it.

In an atomistic detail model, r^N = (r1; r2; ….r_N) represents the complete set of 3N atomic coordinates, being N the number of atoms of the model, while in accord to the statistical mechanic the set of 3N velocities are also required to define the complete microstate of the system.

Non-bonded Interactions

Describing macromolecule and/or supra-molecular systems two main non bonded interaction assume particular importance: the “Van der Waals” forces and the electrostatic forces, the former are generated between the external part of the electronic density of two approaching molecules, and are considered as short range forces. Electrostatic interaction are supposed to be applied between partial charges of polar groups, for example (δ+)H-O(δ-), and

also between groups owning finite charges, like -COO(-) and –NH3(+) or other ionizable groups.

The Lennard-Jones potential is the most commonly used form to describe empirically the van der waals forces:

$$v^{LJ}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \quad (2)$$

with two parameters : σ , the diameter, and Ɛ, the well depth.

If electrostatic charges are present, we add the appropriate Coulomb potentials

$$v^{Columb}(r) = \frac{Q_1 Q_2}{4\pi\epsilon_0 r}, \qquad (3)$$

where Q_1, Q_2 are the charges and ϵ_0 is the permittivity of free space.

Bonding Potentials

For molecular systems, we simply build the molecules out of site-site potentials of the form of Eq.(2) or similar. Typically, a single-molecule quantum-chemical calculation may be applied to estimate the electron density throughout the molecule, and this may then be modeled by a distribution of partial charges via Eq. (3), or more accurately by a distribution of electrostatic multipoles[91]. For molecules, additionally, the intra-molecular bonding interaction must be considered. The simplest molecular model will include terms as reported below:

$$\mathcal{U}_{intramolecular} = \frac{1}{2}\sum_{bonds} k_{ij}^{r}\,(r_{ij} - r_{ij}^{eq})^2 \qquad (4a)$$

$$+\frac{1}{2}\sum_{bend\ angles} k_{ijk}^{0}\,(\theta_{ij} - \theta_{ijk}^{eq})^2 \qquad (4b)$$

$$+\frac{1}{2}\sum_{torsion\ angles}\sum k_{ijkl}^{\phi,m}\,(1 + \cos(m\theta_{ijkl} - \gamma_m))(4c)$$

The “bonds” will typically involve the separation $r_{ij} = |r_i - r_j|$ between adjacent pairs of atoms in a molecular framework, and we assume in Eq. (4a)

an harmonic form with specifed equilibrium bond length r_{ij}^{eq} and energy deformation quantified by the "spring constant" k_{ji}^{r} . This functional form allows to describe the bond length degree of freedom as an harmonic oscillator, but in principle more complex functional form including an higher number of parameters is possible. The "bond angles" θ_{ijk} are between successive bond vectors such as r_i-r_j and r_j-r_k, and therefore involve three atom coordinates:

$$\cos\theta_{ijk} = \hat{r}_{ij} \cdot \hat{r}_{ik} = (\hat{r}_{ij}\hat{r}_{ij})^{-\frac{1}{2}}(\hat{r}_{jk}\hat{r}_{jk})^{-\frac{1}{2}}(\hat{r}_{ij} \cdot \hat{r}_{jk}) \qquad (5)$$

where $\hat{r} = \frac{r}{r}$.

Usually, this bending term is taken to be quadratic in the angular displacement from the equilibrium value, as in Eq. (4b), although periodic functions are also used. Even in that case θ_{ijk}^{eq} and k_{ijk} are the equilibrium bond angle and the corresponding "bending stiffness". The "torsion angles" are defined in terms of three connected bonds, hence four atomic coordinates:

$\cos\phi_{ijkl} = -\hat{n}_{ijk} \cdot \hat{n}_{jkl}$, where $n_{ijk} = r_{ij} \times r_{jk}$, $n_{jkl} = r_{jk} \times r_{kl}$, and $\hat{n} = \frac{n}{n}$, the unit normal to the plane defined by each pair of bonds. Tipically the torsional potential involves an expansion in periodic functions of order m=1,2,…, Eq. (4c). A simulation package force-field will specify the precise form of Eq. (4), and the various strength parameters k and other constants therein. Actually, Eq. (4) is a considerable oversimplifcation. Molecular mechanics force-fields, aimed at accurately predicting structures and properties, will include many cross-terms (e.g. stretch-bend). Quantum mechanical calculations may give a guide to the "best" molecular force-field; also comparison of simulation results with thermophysical properties and vibration frequencies is invaluable in force-field development and refinement. A separate family of force fields, such as

AMBER[92][93], CHARMM, are oriented to biological macromolecule (proteins, polymers) in condensed phases; their functional form is similar to Eq. (4), and their parameters are typically determined by quantum chemical calculations combined with thermo-physical and phase coexistence data.

Force Calculation

Having specified the potential energy function $\mathrm{U}(\mathrm{r}^{\mathrm{N}})$, the next step is to calculate the atomic forces

$$f_i = -\frac{\vartheta}{\vartheta r_i}\mathcal{U}(r^N) \qquad (6)$$

It is quite common practice in classical computer simulations not to attempt to represent intramolecular bonds by terms in the potential energy function, because these bonds have very high vibration frequencies (and arguably should be treated in a quantum mechanical way rather than in the classical approximation). Instead, the bonds are treated as being constrained to have fixed length.

REFERENCES

[1] S. Elli, E. Macchi, T. R. Rudd, R. Raman, G. Sassaki, K. Viswanathan, E. a Yates, Z. Shriver, A. Naggi, G. Torri, R. Sasisekharan, and M. Guerrini, "Insights into the human glycan receptor conformation of 1918 pandemic hemagglutinin-glycan complexes derived from nuclear magnetic resonance and molecular dynamics studies.," *Biochemistry*, vol. 53, no. 25, pp. 4122–35, Jul. 2014.

[2] L. Simonsen, D. R. Olson, C. Viboud, E. Heiman, R. J. Taylor, M. A. Miller, and Reichert, *PANDEMIC INFLUENZA AND MORTALITY: PAST EVIDENCE AND PROJECTIONS FOR THE FUTURE*. The National Academies Press, 2004, pp. 89–114.

[3] D. J. Smith, A. S. Lapedes, J. C. de Jong, T. M. Bestebroer, G. F. Rimmelzwaan, A. D. M. E. Osterhaus, and R. a M. Fouchier, "Mapping the antigenic and genetic evolution of influenza virus.," *Science*, vol. 305, no. 5682, pp. 371–6, Jul. 2004.

[4] J. B. Plotkin, J. Dushoff, and S. a Levin, "Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 9, pp. 6263–8, Apr. 2002.

[5] G. W. Both, M. J. Sleigh, N. J. Cox, and a P. Kendal, "Antigenic drift in influenza virus H3 hemagglutinin from 1968 to 1980: multiple evolutionary pathways and sequential amino acid changes at key antigenic sites.," *J. Virol.*, vol. 48, no. 1, pp. 52–60, Oct. 1983.

[6] F. Carrat and a Flahault, "Influenza vaccine: the challenge of antigenic drift.," *Vaccine*, vol. 25, no. 39–40, pp. 6852–62, Sep. 2007.

[7] Bernd Sebastian Kamps, C. Hoffmann, and W. Preiser, *Influenza report 2006*. Flying publisher, 2006.

[8] R. J. Webby and R. G. Webster, "Emergence of influenza A viruses.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 356, no. 1416, pp. 1817–28, Dec. 2001.

[9] J. Treanor, "Influenza vaccine--outmaneuvering antigenic shift and drift.," *N. Engl. J. Med.*, vol. 350, no. 3, pp. 218–20, Jan. 2004.

[10] K. Das, J. M. Aramini, L.-C. Ma, R. M. Krug, and E. Arnold, "Structures of influenza A proteins and insights into antiviral drug targets.," *Nat. Struct. Mol. Biol.*, vol. 17, no. 5, pp. 530–8, May 2010.

[11] S. Yamada, Y. Suzuki, T. Suzuki, M. Q. Le, C. a Nidom, Y. Sakai-Tagawa, Y. Muramoto, M. Ito, M. Kiso, T. Horimoto, K. Shinya, T. Sawada, M. Kiso, T. Usui, T. Murata, Y. Lin, A. Hay,

L. F. Haire, D. J. Stevens, R. J. Russell, S. J. Gamblin, J. J. Skehel, and Y. Kawaoka, “Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors.,” *Nature*, vol. 444, no. 7117, pp. 378–82, Nov. 2006.

[12] J. J. Skehel and D. C. Wiley, “Receptor binding and membrane fusion in virus entry: The Influenza Hemagglutinin,” 2000.

[13] Q. Huang, “Early steps of the conformational change of influenza virus hemagglutinin to a fusion active state Stability and energetics of the hemagglutinin,” *Biochim. Biophys. Acta - Biomembr.*, vol. 1614, no. 1, pp. 3–13, Jul. 2003.

[14] W. W. Wu and N. Panté, “The directionality of the nuclear transport of the influenza A genome is driven by selective exposure of nuclear localization sequences on nucleoprotein.,” *Virol. J.*, vol. 6, p. 68, Jan. 2009.

[15] L. L. Newcomb, R.-L. Kuo, Q. Ye, Y. Jiang, Y. J. Tao, and R. M. Krug, “Interaction of the influenza a virus nucleocapsid protein with the viral RNA polymerase potentiates unprimed viral RNA replication.,” *J. Virol.*, vol. 83, no. 1, pp. 29–36, Jan. 2009.

[16] S. J. Plotch, M. Bouloy, and R. M. Krug, “A Unique Cap (m7GpppXm) -Dependent Influenza Virion Endonuclease Cleaves Capped RNAs to Generate the Primers That Initiate Viral RNA Transcription,” vol. 23, no. March, pp. 847–858, 1981.

[17] T. Samji, “Influenza A: understanding the viral life cycle.,” *Yale J. Biol. Med.*, vol. 82, no. 4, pp. 153–9, Dec. 2009.

[18] S. J. Plotch, M. Bouloy, I. Ulmanen, and R. M. Krug, “A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription.,” *Cell*, vol. 23, no. 3, pp. 847–58, Mar. 1981.

[19] M. Hagen, T. D. Chung, J. a Butcher, and M. Krystal, “Recombinant influenza virus polymerase: requirement of both 5’ and 3' viral ends for endonuclease activity.,” *J. Virol.*, vol. 68, no. 3, pp. 1509–15, Mar. 1994.

[20] M. E. Nemeroff, S. M. Barabino, Y. Li, W. Keller, and R. M. Krug, “Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3’end formation of cellular pre-mRNAs.,” *Mol. Cell*, vol. 1, no. 7, pp. 991–1000, Jun. 1998.

[21] K. Shimizu, A. Iguchi, R. Gomyou, and Y. Ono, “Influenza virus inhibits cleavage of the HSP70 pre-mRNAs at the polyadenylation site.,” *Virology*, vol. 254, no. 2, pp. 213–9, Feb. 1999.

[22] G. Neumann, M. T. Hughes, and Y. Kawaoka, “Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1.,” *EMBO J.*, vol. 19, no. 24, pp. 6751–8, Dec. 2000.

[23] D. P. Nayak, E. K.-W. Hui, and S. Barman, “Assembly and budding of influenza virus.,” *Virus Res.*, vol. 106, no. 2, pp. 147–65, Dec. 2004.

[24] R. G. Webster, W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka, “Evolution and ecology of influenza A viruses.,” *Microbiol. Rev.*, vol. 56, no. 1, pp. 152–79, Mar. 1992.

[25] D. J. Alexander, “A review of avian influenza in different bird species.,” *Veterinary microbiology*, vol. 74, no. 1–2. pp. 3–13, 22-May-2000.

[26] L. E. L. Perkins and D. E. Swayne, “Comparative Susceptibility of Selected Avian and Mammalian Species to a Hong Kong–Origin H5N1 High-Pathogenicity Avian Influenza Virus,” *Avian Dis.*, vol. 47, no. s3, pp. 956–967, Sep. 2003.

[27] J. K. Taubenberger, A. H. Reid, R. M. Lourens, R. Wang, G. Jin, and T. G. Fanning, “Characterization of the 1918 influenza virus polymerase genes.,” *Nature*, vol. 437, no. 7060, pp. 889–93, Oct. 2005.

[28] N. P. a. S. Johnson and J. Mueller, “Updating the Accounts: Global Mortality of the 1918-1920 ‘Spanish’ Influenza Pandemic,” *Bull. Hist. Med.*, vol. 76, no. 1, pp. 105–115, 2002.

[29] a H. Reid, T. G. Fanning, J. V Hultin, and J. K. Taubenberger, “Origin and evolution of the 1918 ‘Spanish’ influenza virus hemagglutinin gene.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 4, pp. 1651–6, Feb. 1999.

[30] J. K. Thomas and J. Noppenberger, “Avian influenza: a review.,” *Am. J. Health. Syst. Pharm.*, vol. 64, no. 2, pp. 149–65, Jan. 2007.

[31] R. B. Belshe, “The origins of pandemic influenza--lessons from the 1918 virus.,” *N. Engl. J. Med.*, vol. 353, no. 21, pp. 2209–11, Nov. 2005.

[32] Ned Rozell, “Permafrost Preserves Clues to Deadly 1918 Flu, Alaska Science Forum,” *Alaska Science Forum*, 1998. [Online]. Available: http://www2.gi.alaska.edu/ScienceForum/ASF13/1386.html.

[33] W. H. Organization, *Avian influenza: assessing the pandemic threat*. 2005.

[34] W. H. Organization, “Evolution of a pandemic A(H1N1) 2009,” 2010.

[35] Y. Shi, W. Zhang, F. Wang, J. Qi, Y. Wu, H. Song, F. Gao, Y. Bi, Y. Zhang, Z. Fan, C. Qin, H. Sun, J. Liu, J. Haywood, W. Liu, W. Gong, D. Wang, Y. Shu, Y. Wang, J. Yan, and G. F. Gao, "Structures and receptor binding of hemagglutinins from human-infecting H7N9 influenza viruses.," *Science*, vol. 342, no. 6155, pp. 243–7, Oct. 2013.

[36] WHO, "Background and summary of human infection with avian influenza A(H7N9) virus – as of 31 January 2014," 2014. [Online]. Available: http://www.who.int/influenza/human_animal_interface/20140131_background_and_summary_H7N9_v1.pdf?ua=1.

[37] T. Watanabe, S. Watanabe, E. a Maher, G. Neumann, and Y. Kawaoka, "Pandemic potential of avian influenza A (H7N9) viruses.," *Trends Microbiol.*, vol. 22, no. 11, pp. 623–631, Sep. 2014.

[38] J. Zhou, D. Wang, R. Gao, B. Zhao, J. Song, X. Qi, Y. Zhang, Y. Shi, L. Yang, W. Zhu, T. Bai, K. Qin, Y. Lan, S. Zou, J. Guo, J. Dong, L. Dong, Y. Zhang, H. Wei, X. Li, J. Lu, L. Liu, X. Zhao, X. Li, W. Huang, L. Wen, H. Bo, L. Xin, Y. Chen, C. Xu, Y. Pei, Y. Yang, X. Zhang, S. Wang, Z. Feng, J. Han, W. Yang, G. F. Gao, G. Wu, D. Li, Y. Wang, and Y. Shu, "Biological features of novel avian influenza A (H7N9) virus.," *Nature*, vol. 499, no. 7459, pp. 500–3, Jul. 2013.

[39] T. Watanabe, M. Kiso, S. Fukuyama, N. Nakajima, M. Imai, S. Yamada, S. Murakami, S. Yamayoshi, K. Iwatsuki-Horimoto, Y. Sakoda, E. Takashita, R. McBride, T. Noda, M. Hatta, H. Imai, D. Zhao, N. Kishida, M. Shirakura, R. P. de Vries, S. Shichinohe, M. Okamatsu, T. Tamura, Y. Tomita, N. Fujimoto, K. Goto, H. Katsura, E. Kawakami, I. Ishikawa, S. Watanabe, M. Ito, Y. Sakai-Tagawa, Y. Sugita, R. Uraki, R. Yamaji, A. J. Eisfeld, G. Zhong, S. Fan, J. Ping, E. a Maher, A. Hanson, Y. Uchida, T. Saito, M. Ozawa, G. Neumann, H. Kida, T. Odagiri, J. C. Paulson, H. Hasegawa, M. Tashiro, and Y. Kawaoka, "Characterization of H7N9 influenza A viruses isolated from humans.," *Nature*, vol. 501, no. 7468, pp. 551–5, Sep. 2013.

[40] J. Liu, H. Xiao, Y. Wu, D. Liu, X. Qi, Y. Shi, and G. F. Gao, "H7N9: a low pathogenic avian influenza A virus infecting humans.," *Curr. Opin. Virol.*, vol. 5, no. February 2013, pp. 91–7, Apr. 2014.

[41] Q. Zhang, J. Shi, G. Deng, J. Guo, X. Zeng, X. He, H. Kong, C. Gu, X. Li, J. Liu, G. Wang, Y. Chen, L. Liu, L. Liang, Y. Li, J. Fan, J. Wang, W. Li, L. Guan, Q. Li, H. Yang, P. Chen, L. Jiang, Y. Guan, X. Xin, Y. Jiang, G. Tian, X. Wang, C. Qiao, C. Li, Z. Bu, and H. Chen, "H7N9 influenza viruses are transmissible in ferrets by respiratory droplet.," *Science*, vol. 341, no. 6144, pp. 410–4, Jul. 2013.

[42] X. Xiong, S. R. Martin, L. F. Haire, S. a Wharton, R. S. Daniels, M. S. Bennett, J. W. McCauley, P. J. Collins, P. a Walker, J. J. Skehel, and S. J. Gamblin, "Receptor binding by an H7N9 influenza virus from humans.," *Nature*, vol. 499, no. 7459, pp. 496–9, Jul. 2013.

[43] M. Richard, E. J. a Schrauwen, M. de Graaf, T. M. Bestebroer, M. I. J. Spronken, S. van Boheemen, D. de Meulder, P. Lexmond, M. Linster, S. Herfst, D. J. Smith, J. M. van den Brand, D. F. Burke, T. Kuiken, G. F. Rimmelzwaan, A. D. M. E. Osterhaus, and R. a M. Fouchier, "Limited airborne transmission of H7N9 influenza A virus between ferrets.," *Nature*, vol. 501, no. 7468, pp. 560–3, Sep. 2013.

[44] K. Tharakaraman, A. Jayaraman, R. Raman, K. Viswanathan, N. W. Stebbins, D. Johnson, Z. Shriver, V. Sasisekharan, and R. Sasisekharan, "Glycan receptor binding of the influenza A virus H7N9 hemagglutinin.," *Cell*, vol. 153, no. 7, pp. 1486–93, Jun. 2013.

[45] R. J. Russell, D. J. Stevens, L. F. Haire, S. J. Gamblin, and J. J. Skehel, "Avian and human receptor binding by hemagglutinins of influenza A viruses.," *Glycoconj. J.*, vol. 23, no. 1–2, pp. 85–92, Feb. 2006.

[46] A. Chandrasekaran, A. Srinivasan, R. Raman, K. Viswanathan, S. Raguram, T. M. Tumpey, V. Sasisekharan, and R. Sasisekharan, "Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin.," *Nat. Biotechnol.*, vol. 26, no. 1, pp. 107–13, Jan. 2008.

[47] D. Xu, E. I. Newhouse, R. E. Amaro, H. C. Pao, L. S. Cheng, P. R. L. Markwick, J. A. McCammon, W. W. Li, and P. W. Arzberger, "Distinct glycan topology for avian and human sialopentasaccharide receptor analogues upon binding different hemagglutinins: a molecular dynamics perspective.," *J. Mol. Biol.*, vol. 387, no. 2, pp. 465–91, Mar. 2009.

[48] T. M. Tumpey, T. R. Maines, N. Van Hoeven, L. Glaser, A. Solórzano, C. Pappas, N. J. Cox, D. E. Swayne, P. Palese, J. M. Katz, and A. García-Sastre, "A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission.," *Science*, vol. 315, no. 5812, pp. 655–9, Feb. 2007.

[49] G. Neumann, C. a Macken, and Y. Kawaoka, "Identification of amino acid changes that may have been critical for the genesis of A(H7N9) influenza viruses.," *J. Virol.*, vol. 88, no. 9, pp. 4877–96, May 2014.

[50] T. Ito, J. N. S. S. Couceiro, S. Kelm, G. Baum, S. Krauss, M. R. Castrucci, H. Kida, J. C. Paulson, G. Robert, Y. Kawaoka, L. G. Baum, and R. G. Webster, "Molecular Basis for the Generation in Pigs of Influenza A Viruses with Pandemic Potential Molecular Basis for the Generation in Pigs of Influenza A Viruses with Pandemic Potential," *J. Virol.*, vol. 72, no. 9, pp. 7367–7373, 1998.

[51] M. N. Matrosovich, S. Gambaryan, S. Teneberg, V. E. Piskarev, S. S. Yamnikova, D. K. Lvov, J. S. Robertson, and K. a Karlsson, "Avian influenza A viruses differ from human viruses by recognition of sialyloligosaccharides and gangliosides and by a higher conservation of the HA receptor-binding site.," *Virology*, vol. 233, no. 1, pp. 224–34, Jun. 1997.

[52] S. Gambaryan, B. Tuzikov, V. E. Piskarev, S. S. Yamnikova, D. K. Lvov, J. S. Robertson, N. V Bovin, and M. N. Matrosovich, "Specification of receptor-binding phenotypes of influenza virus isolates from different hosts using synthetic sialylglycopolymers: non-egg-adapted human H1 and H3 influenza A and influenza B viruses share a common high binding affinity for 6'-sialyl(N-ace," *Virology*, vol. 232, no. 2, pp. 345–50, Jun. 1997.

[53] T. J. Pritchetts and C. Paulson, "Basis for the Potent Inhibition of Influenza Virus Infection by Equine and Guinea Pig alfa2-Macroglobulin," *J. Biol. Chem.*, vol. 264, no. 17, pp. 9850–9858, 1989.

[54] W. D. C. Rogers G.N., Paulson J.C., Daniels R.S., Skehel J.J., Wilson J.A., "Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity," *Nature*, vol. 173, pp. 76–78, 1983.

[55] G. Neumann, T. Noda, and Y. Kawaoka, "Emergence and pandemic potential of swine-origin H1N1 influenza virus.," *Nature*, vol. 459, no. 7249, pp. 931–9, Jun. 2009.

[56] J. Stevens, O. Blixt, T. M. Tumpey, J. K. Taubenberger, J. C. Paulson, and I. a Wilson, "Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus.," *Science*, vol. 312, no. 5772, pp. 404–10, Apr. 2006.

[57] Y. Ha, D. J. Stevens, J. J. Skehel, and D. C. Wiley, "X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 20, pp. 11181–6, Sep. 2001.

[58] S. J. Gamblin, L. F. Haire, R. J. Russell, and D. J. Stevens, "The Structure and Receptor Binding Properties of the 1918," vol. 303, no. March, pp. 1838–1842, 2004.

[59] J. Liu, D. J. Stevens, L. F. Haire, P. a Walker, P. J. Coombs, R. J. Russell, S. J. Gamblin, and J. J. Skehel, "Structures of receptor complexes formed by hemagglutinins from the Asian Influenza pandemic of 1957.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 40, pp. 17175–80, Oct. 2009.

[60] Y. Ha, D. J. Stevens, J. J. Skehel, and D. C. Wiley, "X-ray structure of the hemagglutinin of a potential H3 avian progenitor of the 1968 Hong Kong pandemic influenza virus☆," *Virology*, vol. 309, no. 2, pp. 209–218, May 2003.

[61] R. Xu, R. McBride, J. C. Paulson, C. F. Basler, and I. a Wilson, "Structure, receptor binding, and antigenicity of influenza virus hemagglutinins from the 1957 H2N2 pandemic.," *J. Virol.*, vol. 84, no. 4, pp. 1715–21, Feb. 2010.

[62] M. Basner, D. F. Dinges, D. Mollicone, C. W. Jones, E. C. Hyder, A. Di Antonio, I. Savelev, K. Kan, N. Goel, and B. V Morukov, "Evolution of the receptor binding properties of the

influenza A(H3N2) hemagglutinin," *Proc. Natl. Acad. Sci.*, vol. 110, no. 7, pp. 2676–2676, Jan. 2013.

[63] A. H. Kwan, M. Mobli, P. R. Gooley, G. F. King, and J. P. Mackay, "Macromolecular NMR spectroscopy for the non-spectroscopist.," *FEBS J.*, vol. 278, no. 5, pp. 687–703, Mar. 2011.

[64] J. H. F. Bothwell and J. L. Griffin, "An introduction to biological nuclear magnetic resonance spectroscopy," *Biol. Rev.*, vol. 86, no. 2, pp. 493–510, May 2011.

[65] T. L. James, "Chapter 1 Fundamentals of NMR," pp. 1–31, 1998.

[66] T. D. W. Claridge, *High-Resolution NMR Techniques in Organic Chemistry*, Tetrahedro. 2009.

[67] "Macromolecular NMR spectroscopy for the non-spectroscopist," University course .

[68] H. Friebolin, *Basic One and Two-Dimensional Spectroscopy*. Wiley-VCH, 2010.

[69] C. a Lepre, J. M. Moore, and J. W. Peng, "Theory and applications of NMR-based screening in pharmaceutical research.," *Chem. Rev.*, vol. 104, no. 8, pp. 3641–76, Aug. 2004.

[70] J. de P. Peter Atkins, *Atkins - Physical Chemistry 8th Edition.pdf*, Oxford Pre. 2006.

[71] E. P. Igor Tvaroska, Milos Hricovini, "An attempt to derive a new karplus type equation of vicinal proton carbon coupling constant for C-O-C-H segments of bonded atoms," *Carbohydr. Res.*, vol. 189, pp. 359–362, 1989.

[72] K. W. and R. R. E. C.Griesinger, G. Otting, "Clean TOCSY for 1H Spin System Identification in Macromolecules," *J. Magn. Reson.*, vol. 110, no. 13, pp. 7870–7872, 1988.

[73] K. P. and B. A. Arepalli, S. R., Glaudemans C. P. J., Daves G. D., "Identification of protein-mediated indirect NOE effects in a disaccharide-Fab' complex by transferred ROESY.pdf," *J.Magn.Reson.Ser.B*, vol. 106, pp. 195–198, 1995.

[74] I. P. Gerothanassis, A. Troganis, V. Exarchou, and K. Barbarossou, "Nuclear Magnetic Resonance (Nmr) Spectroscopy: Basic Principles and Phenomena, and Their Applications To Chemistry, Biology and Medicine," *Chem. Educ. Res. Pract.*, vol. 3, no. 2, p. 229, 2002.

[75] H. Friebolin, *1 The Physical Basis of NMR Spectroscopy*. 2010.

[76] S. Berger, “NMR techniques employing selective radiofrequency pulses in combination with pulsed field gradients,” *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 30, no. 3–4, pp. 137–156, Jul. 1997.

[77] A. Meissner and O. W. Sørensen, “Measurement of J (H , H) and long-range J (X , H) coupling constants in small molecules . Broadband XLOC and J-HMBC,” pp. 49–52, 2001.

[78] M. D. Bax Ad, “Improved Resolution and Sensitivity in ‘H-Detected Heteronuclear Multiple-Bond Correlation Spectroscopy,” *J. Magn. Reson.*, vol. 78, pp. 186–191, 1988.

[79] C. E. Hadden, G. E. Martin, and V. V Krishnamurthy, “Improved performance accordion heteronuclear multiple-bond correlation spectroscopy-IMPEACH-MBC.,” *J. Magn. Reson.*, vol. 140, no. 1, pp. 274–80, Sep. 1999.

[80] L. P. Calle, F. J. Cañada, and J. Jiménez-Barbero, “Application of NMR methods to the study of the interaction of natural products with biomolecular receptors.,” *Nat. Prod. Rep.*, vol. 28, no. 6, pp. 1118–25, Jun. 2011.

[81] J. W. Peng, J. Moore, and N. Abdul-Manan, “NMR experiments for lead generation in drug discovery,” *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 44, no. 3–4, pp. 225–256, Jul. 2004.

[82] B. Meyer and T. Peters, “NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors,” *Angew. Chem. Int. Ed. Engl.*, vol. 42, no. 8, pp. 864–890, 2003.

[83] F. Ni, “Recent developments in transferred noe methods,” *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 26, pp. 517–606, 1994.

[84] A. Chen and M. J. Shapiro, “NOE Pumping. A High-Throughput Method To Determine Compounds with Binding Affinity to Macromolecules by NMR,” *J. Am. Chem. Soc.*, vol. 122, no. 2, pp. 414–415, Jan. 2000.

[85] M. Lin, M. J. Shapiro, and J. R. Wareing, “Diffusion-Edited NMR Affinity NMR for Direct Observation of Molecular Interactions,” *J. Am. Chem. Soc.*, vol. 7863, no. 6, pp. 5249–5250, 1997.

[86] T. Carlomagno, “Ligand-target interactions: what can we learn from NMR?,” *Annu. Rev. Biophys. Biomol. Struct.*, vol. 34, pp. 245–66, Jan. 2005.

[87] L. Fielding, “NMR methods for the determination of protein–ligand dissociation constants,” *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 51, no. 4, pp. 219–242, Nov. 2007.

[88] B. Applications and N. M. R. Spectroscopy, “Macromolecular Complexes : Structures and Interactions by NMR,” in *Applications, Biochemical Spectroscopy, N M R*, 2010.

[89] M. Pellecchia, “Solution nuclear magnetic resonance spectroscopy techniques for probing intermolecular interactions.,” *Chem. Biol.*, vol. 12, no. 9, pp. 961–71, Sep. 2005.

[90] M. Mayer and B. Meyer, “Mapping the Active Site of Angiotensin-Converting Enzyme by Transferred NOE Spectroscopy,” *J. Med. Chem.*, vol. 43, no. 11, pp. 2093–2099, Jun. 2000.

[91] M. P. Allen, “Introduction to Molecular Dynamics Simulation,” in *Computational Soft Matter: From Synthetic Polymers to Proteins.*, vol. 23, 2004, pp. 1–28.

[92] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weinerl, “A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins,” *J. Am. Chem. Soc.*, vol. 106, no. 17, pp. 765–784, 1984.

[93] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. a. Kollman, “A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules,” *J. Am. Chem. Soc.*, vol. 117, no. 19, pp. 5179–5197, May 1995.

Results and Discussion

Objective of the thesis

As previously described in the introduction, Influenza A viruses have been viewed as 'unpredictable' pathogens, because they have the potentiality to lead to a widespread pandemic outbreak acquiring the ability to bind with high affinity human receptor. One of the most relevant aspects governing virus selectivity is the interaction between the influenza virus surface protein hemagglutinin (HA) and its cell surface glycan receptors, which present the same composition but they differ for the type of the linkage between the Neu5Ac and the Gal residues, which is α(1-3) in the avian receptor, whereas is α(1-6) in the human receptor.

Co-crystal structures of the receptors bound to HA have been previously described, highlighting differences in glycan receptor conformation. As a consequence, the conformation and topology characterization of the human and avian sialylated glycan receptors in solution is required, given that it is plausible to presume that receptor dynamic properties, strictly related to its structure, could play a role in the recognition and binding of HA proteins.

In order to obtain the fully description of the conformational properties of the two sialylated pentasaccharides (LSTc, Neu5Ac-α-(2-6)-Gal-β-(1-4)-GlcNAc-β-(1-3)-Gal-β-(1-4)-Glc; and LSTa, Neu5Ac-α-(2-3)-Gal-β-(1-3)-GlcNAc-β-(1-3)-Gal-β-(1-4)-Glc), used as models for human and avian receptors, extensive NMR analysis and MD simulation, followed by the numerical analysis of the data generated, were used.

The second step of our work was the interaction studies, based on NMR techniques, between selected viral HAs and human and avian glycan cell

receptor models. The chosen hemagglutinins were SC18-HA (hemagglutinin coming from the virus A/South Carolina/1/1918, H1N1, responsible of the pandemic event in 1918) and two of mutant versions: a single (Asp225-Gly) and double mutant (Asp225-Gly and Asp190Glu), that were proved to switch their affinity from LSTc (human) to LSTa (avian) receptors. The aim of this study was to capture the active changes in conformation of LSTa and LSTc induced upon binding to HA and sheds light on distinct structural constraints imposed by the RBS of the wild type and mutated versions of the virus.

We used MD simulation to generate models of our systems, allowing the structural and dynamic comparison between the different complexes, and obtaining together with NMR a complete overview of these interactions.

Finally, we used NMR and MD simulation also to characterize the new and less pandemic virus (H7N9) that shows the ability to infect both birds and humans, but it is unable to diffuse between humans. The complexes of avian and human receptors with the H7 wild type virus, isolated from Anhui H7N9 virus in 2013, and its mutant forms (Gly228Ser and Gly228Ser/Val186Gly), that are supposed to have a higher affinity toward the human receptor, have been characterized.

Understanding the structural basis of the avian virus hemagglutinin interaction with glycans, that affects its specificity toward avian or human cell receptor, will be very helpful for surveillance of emerging new influenza virus potentially pandemic for human species.

Human (α2-6) and avian (α2-3) sialylated receptors of influenza A virus show distinct conformations and dynamics in solution[1]

INTRODUCTION

The interaction between the influenza A virus surface protein haemagglutinin (HA) and its cell surface glycan receptors is one of the critical steps governing virus selectivity. Several crystal structures of different HAs have been solved, both native and co-crystallised with glycan receptors[2]–[6]. These studies have provided an insight into the ability of HA to interact with receptor glycans. Human-adapted influenza viruses preferentially bind to sialylated glycan receptors terminated by N-acetyl neuraminic acid (Neu5Ac) α2->6 linked to galactose (Gal) (human receptors), which are predominantly expressed on the apical surface of the upper respiratory epithelia in humans [7]–[9]. In contrast, avian-adapted influenza viruses preferentially bind to sialylated glycans terminated by Neu5Ac α2->3 linked to Gal, which are expressed predominantly in the gut and respiratory tract of birds and, importantly, in the deep lung (alveolar region) of humans[8]. Therefore, one of the characteristics of human-adaptation of influenza A viruses is the change in glycan receptor binding preference of HA from avian to human receptors[10]–[15].

Recent studies based on analysis of HA-glycan co-crystal structures[16][17] have highlighted the importance of glycan conformation and topology in governing receptor specificity of avian- and human-adapted HAs. In the case of avian receptors, the majority of contacts between HA and glycan involves the

Neu5Ac residue at the non-reducing end. Chandrasekaran *et al*. [17]proposed that, by fixing the Neu5Ac in the glycan-binding pocket of HA, the other residues in this trisaccharide motif sample different torsional angles at their glycosidic linkages, such that the space occupied by the whole glycan resembles a *cone* with Neu5Ac at the vertex, characterised by a topological angle $\theta > 100^{\circ}$ (Fig. 10 of Introduction). The topological angle θ is defined by the anomeric carbon of the three residues Neu5Ac, Gal and GlcNAc (N-acetyl glucosamine) in going from the non-reducing to the reducing end. They also proposed for the case of human receptors with a long enough glycan motif, comprising at least a tetrasaccharide Neu5Ac-α(2→6)-Gal-β(1→4)-GlcNAc-β(1→3)-Gal- at the non-reducing end, spans the glycan binding surface on HA. The additional C5-C6 bond of the Neu5Ac-α(2→6)-Gal glycosidic linkage provides added mobility to its glycosidic junction and the remaining backbone allowing it to explore conformations similar to an *umbrella* progressing from a fully open to a fully closed form (defined as having a θ angle $< 100^{\circ}$) (Fig. 10 Introduction)[17]. It is clear that the *cone* and *umbrella* topologies are generated by their motions while they are bound to the receptor, instead their conformation in no-bound state appear strongly dependent from their respective non-reducing end connectivity, α(2→3) or α(2→6). The role of glycan shape and topology in regulating binding specificity for HA has also been investigated using molecular dynamics simulations[18], indicating that there are significant differences in the conformational sampling of avian and human receptors from the different subtypes (H3, H5 and H9) upon binding to HA.

The previously mentioned studies highlighted differences in glycan receptor conformation and structural details when co-crystallized bound to HA,

illustrating the structural specificity of HA-glycan interactions. Consequently, it is logical to suppose that receptor dynamic properties, strictly related to its structure, could play a role in recognition and binding by the HA protein. To characterize human and avian sialylated glycan receptors in term of conformation and topology in solution, detailed NMR analyses, MD simulations, followed by a statistical analysis of the sampled trajectories were performed. Two sialylated pentasaccharides: LSTc (Neu5Ac-α(2→6)-Gal-β(1→4)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc) and LSTa (Neu5Ac-α(2→3)-Gal-β(1→3)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc) were used respectively as a model for human and avian receptor (Fig. 1).

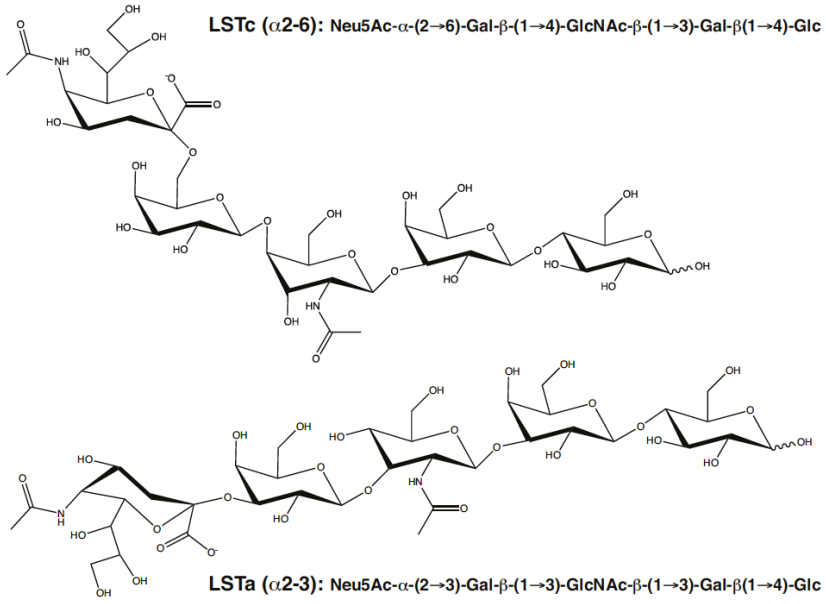


Fig.1 Scheme of the two pentasaccharides used as a model for human, LSTc, and avian, LSTa, receptors.

We refer to Gal1 and Gal2 as labels for the first and second Gal residue encountered in sequence of LSTc and LSTa, from the non-reducing to the reducing end. NOESY, ^{13}C-NMR relaxation measurements and analysis of the glycan MD trajectories, revealed that LSTa and LSTc have distinct conformational dynamics, topology and intra-molecular hydrogen bond networks in solution. Specifically, LSTc preferentially adopts a bent shape in solution, which is consistent with the "umbrella"-like conformation, whereas, LSTa adopts a bow shape in solution, which agrees with a "cone"-like structure. These observations suggest that the different species-adapted HAs could recognise specific solution glycan shape, conformations, and dynamic behaviour.

RESULT AND DISCUSSION

NMR analyses of LSTa and LSTc. NMR characterisation of Sialyl-lacto-*N*-tetraoses LSTa and LSTc (Fig. 1) were carried out using homonuclear (COSY, TOCSY and NOESY), heteronuclear (edited-HSQC and HSQC-TOCSY), and heteronuclear multiple bond correlation HMBC (^{1}H/^{13}C) experiments. This permitted to obtain the complete assignment of the hydrogen and carbon resonances for both LSTc and LSTa. In Fig. 2 the HSQC spectra at 900MHz of LSTc (Top) and LSTa (bottom) are shown, in which is possible to see the correlation between proton and carbon for the pentasaccharides with the correspondent assignment. In Table 1 the corresponding chemical shift assignments for ^{1}H and ^{13}C for both pentasaccharides are reported. TOCSY analysis allowed to observe the connectivity of the pyranosyl rings, with H-H axial correlations confirming the monosaccharide identities.

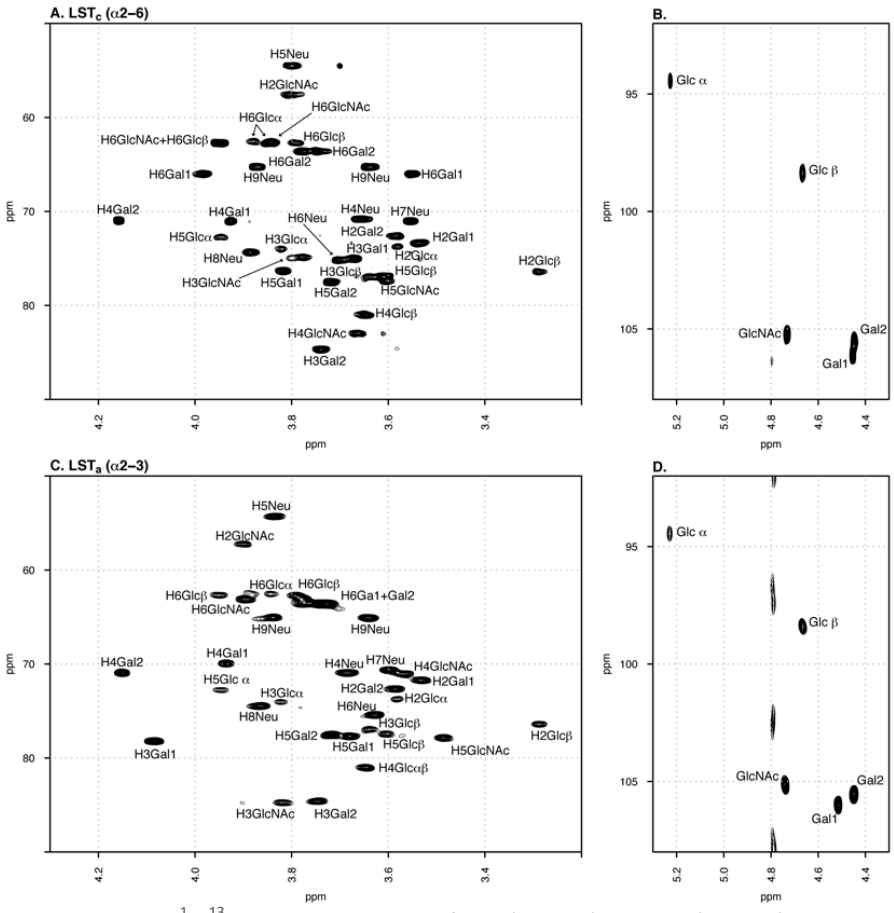


Fig. 2 ^{1}H, ^{13}C HSQC NMR spectra of LSTc (A and B) and LSTa (C and D).

The chemical shifts of the acetamide methyl groups of GlcNAc and Neu5Ac for both LSTa and LSTc were determined by 2D TOCSY (buffered system, 10% D_2O), providing NH correlations with some of the hydrogens of each ring and with those from acetamide (Fig. 3).

			LSTa			
$^{1}H/^{13}C$	NeuAc5	Gal-1	GlcNAc	Gal-2	Glcα	Glcβ
1	176.7	4.513/105.9	4.739/105.1	4.446/105.6	5.227/94.4	4.665/98.3
2	102.3	3.530/71.7	3.898/57.2	3.584/72.6	3.580/73.7	3.256/76.3
3	1.782ax:2.751eq/42.3	4.082/78.2	3.817/84.7	3.745/84.5	3.821/74.0	3.638/76.9
4	3.683/70.9	3.932/69.8	3.565/71.08	4.145/70.9	3.644/81.0	3.644/81.0
5	3.834/54.2	3.677/77.7	3.485/77.8	3.717/77.5	3.944/72.7	3.603/77.4
6	3.627/75.4	3.727^{a};3.771^{b}/63.6	3.892a;3.782 b/63.1	3.727^{a};3.771^{b}/63.6	3.853^{a};3.85^{b}/62.5	3.948^{a};3.841^{b}/62.6
7	3.596/70.6		2.028/24.9			
8	3.862/74.4		177.6			
9	3.836^{a};3.639^{b}/65.0					
10	2.028/24.8					
11	177.6					
NH	8.181 $^{1}J_{H-H}$(9.90)					
NH			8.480 $^{1}J_{H-H}$(10.08)			
			LSTc			
$^{1}H/^{13}C$	NeuAc5	Gal-1	GlcNAc	Gal-2	Glcα	Glcβ
H1	176.2	4.451/105.9	4.73/105.2	4.445/105.5	5.22/94.4	4.664/98.3
H2	102.7	3.534/73.3	3.802/57.3	3.583/72.6	3.58/73.7	3.287/76.4
H3	1.711ax; 2.665eq/42.3	3.671/74.9	3.777/74.9	3.738/84.6	3.820/74.0	3.639/76.94
H4	3.654/70.7	3.926/70.1	3.663/82.9	4.158/70.9	3.649/80.9	3.649/80.9
H5	3.80/54.5	3.818/76.3	3.601/77.3	3.720/77.4	3.947/72.7	3.608/76.94
H6	3.702/75.1	3.983^{a};3.544^{b}/66.0	3.95^{a};3.842^{b}/62.65	3.7726^{a};3.7461^{b}/63.55	3.879^{a};3.842^{b}/62.51	3.952^{a};3.792^{b}/62.58
H7	3.553/71.1		2.050/24.9			
H8	3.887/74.3		177.6			
H9	3.870^{a};3.636^{b}/65.2					
H10	2.025/24.7					
H11	177.6					
NH	8.165 $^{1}J_{H-H}$(9.55)					
NH			8.403 $^{1}J_{H-H}$(9.49)			

Table 1 ^{1}H and ^{13}C assignements of LSTa and LSTc (δ values in ppm at 295K).

However, the strong peak superimposition did not permit the complete hydrogen assignment by homonuclear experiments, these were complemented by HSQC-TOCSY and by the COSY effect generated by homonuclear coupling in sensitivity-enhanced HSQC. These correlation peaks, usually considered as artifacts, are very useful in the assignment of signals in crowded spin systems when performing sensitivity-enhanced HSQC experiments[19] (Fig. 4).

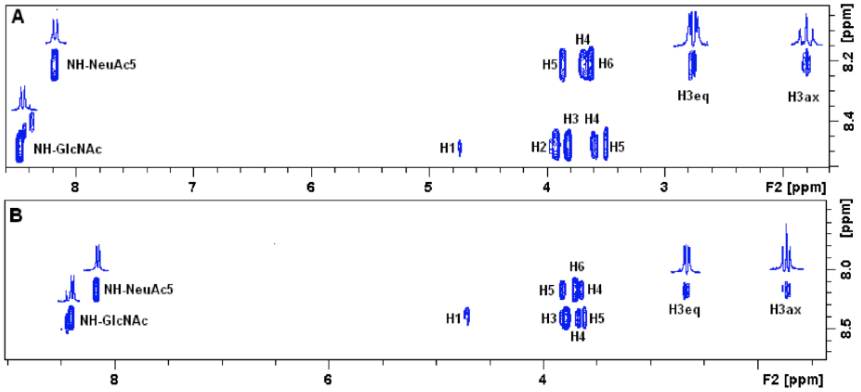


Fig. 3 2D partial TOCSY spectra of NH coupling spin system from Neu5Ac and GlcNAc at 278K, buffered system containing 10% D_2O, pH 7.2. (A) LSTa and (B) LSTc.

Long-range correlations ($^{1}H/^{13}C$) of the quaternary C1 and C2 from the sialyl residues were observed between the H3ax-Neu5Ac at 1.782 ppm with corresponding carbon signals at 176.7 and 102.3 ppm, C1 and C2 respectively.

C2 correlated with H3-Gal1 at 4.082/78.2 ppm of LSTa, confirming the presence of the Neu5Ac α2->3 Gal linkage. Instead, LSTc correlated H3ax-Neu5Ac at 1.711 ppm with carbon signals at 176.2 (C1), 102.7 (C2) ppm, with the latter correlating with H6α/β-Gal1 at 3.983α/3.544β/66.0 ppm, confirming that C2-Neu5Ac is *O*-linked to H6 α/β-Gal1.

Anomeric proton correlations from LSTa were observed at 4.513/84.7 (H1-Gal1/C3-GlcNAc), 4.739/84.5 (H1-GlcNAc /C3-Gal2) and 4.446/81 ppm (H1-Gal2/C4-Glc). LSTc exhibited cross-peaks at 4.451/82.9 (H1-Gal1/C4-GlcNAc), 4.73/84.6 (H1-GlcNAc /C3-Gal2) and 4.445/80.9 ppm (H1-Gal2/C4-Glc). These measurements were confirmed by 1D selective and 2D NOESY experiments.

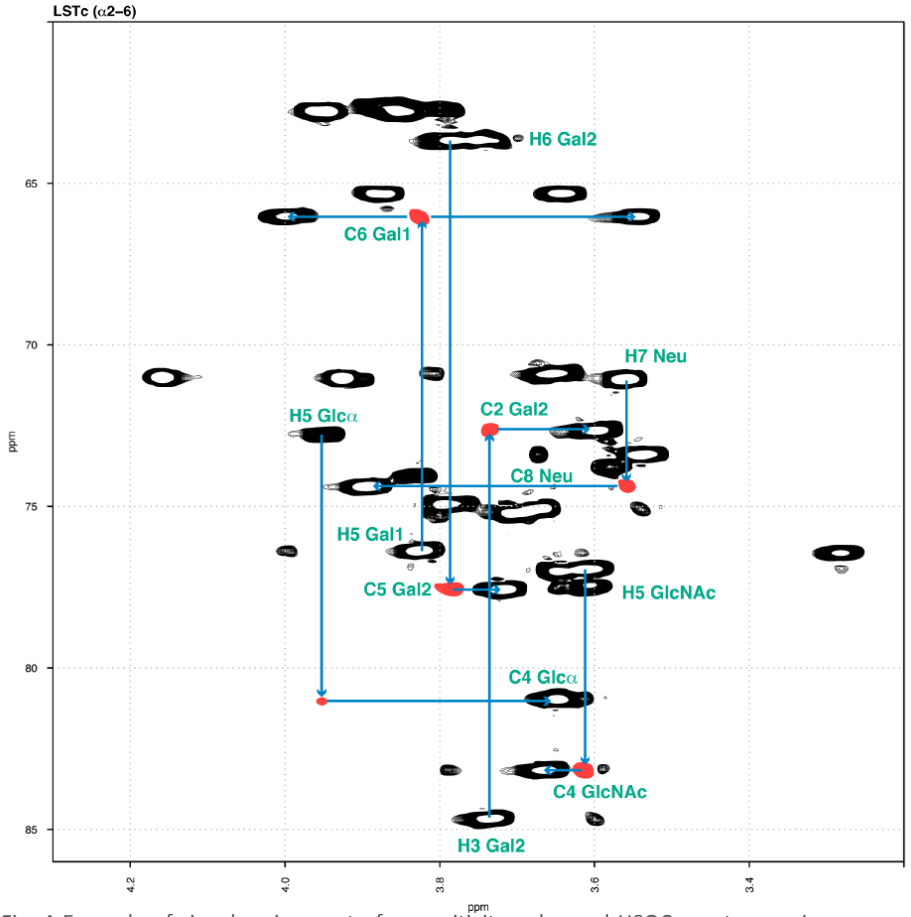


Fig. 4 Example of signal assignment of a sensitivity-enhanced HSQC spectrum using cross peaks generated by homonuclear proton coupling.

Three bond proton-proton coupling constants, $^{3}J_{H\text{-}H}$, were measured for both pentasaccharides, which indicated that all Hex*p* residues assumed the $^{4}C_{1}$ conformation, since $^{3}J_{H1\text{-}H2}$ vary from 7.9-8.2 Hz and $^{3}J_{H3\text{-}H4}$ for Gal*p* residues are ~3.5 Hz, Glc*p* and Glc*p*NAc had values of 9.0 Hz and 9.8 Hz, respectively. The Neu5Ac residue had $^{3}J_{H\text{-}H}$ values compatible with those of a pyranosyl ring in the $^{2}C_{5}$ conformation, $^{3}J_{H3\text{-}H4}$ = $4.8_{(eq)}$ Hz/$12.1_{(ax)}$ Hz and $^{3}J_{H4\text{-}H5}$ = 10.1 Hz, thus

suggesting no significant differences in the ring conformation between the pentasaccharides.

Temperature coefficient (Δppb/K) and CSDs (chemical shifts deviations) were determined by ^{1}H and HSQC experiments but are difficult to interpret, since for both pentasaccharides the shifts are very similar. In order to improve and simplify the differences between them, we used the absolute values of CSD and ^{1}H Δppb/K and converted the value to percentage of shift in each residue (Table 2). Using this approach it becomes easier to observe the temperature effect in ^{1}H and ^{13}C chemical shifts and in global CSD (calculated using the equation 2 in the experimental section).

CSD analysis of LSTc showed that the gradient temperature had a major effect in the reducing end, this behavior is favored by the mutarotation in C1 which provides huge ^{13}C chemical shifts at C4, amplifying the CSD values at least 4 fold when compared to the non-reducing units, similar results were also observed for LSTa. In fact analysis of ^{1}H chemical shift for the reducing ends gave results which were less spread, confirming the amplification effect at C4 produced by mutarotation.

As in proteins, NH temperature coefficients can reflect the extent to which N-acetyl-amide protons are protected from exchange through hydrogen bonding. Evaluation of NH temperature coefficients showed a different behavior between GlcNAc and Neu5Ac. The first had a value of -8.28 ppb/K for LSTc and -8.00 ppb/K for LSTa, while NH for Neu5Ac exhibited lower temperature coefficients with values of -6.77ppb/K and -6.60 ppb/K for LSTc and LSTa, respectively.

$^{1}H/^{13}C$	LSTc (CSD)	LSTa (CSD)	LSTc (ppb/K ^{1}H)	LSTa (ppb/K ^{1}H)	LSTc (ppb/K ^{13}C)	LSTa (ppb/K ^{13}C)	LSTc (CSD) %	LSTa (CSD)%
H3axNeu	0.001068	0.000901	-1.00	-0.26	1.49	3.43	3.1	2.6
H3eqNeu	0.000657	0.000882	0.54	0.18	1.49	3.43	1.9	2.5
H4Neu	0.000857	0.001469	0.65	0.58	-2.22	-5.37	2.5	4.2
H5Neu	0.000935	0.001350	-0.57	-0.53	2.95	4.94	2.7	3.8
H6Neu	0.000789	0.000831	0.29	0.70	2.92	1.78	2.3	2.4
H7Neu	0.001164	0.001195	0.07	-0.04	4.62	4.75	3.4	3.4
H8Neu	0.000454	0.001272	-0.40	-0.86	0.85	-3.73	1.3	3.6
H9Neu	0.001434	0.001466	0.10	0.34	-5.69	5.67	4.2	4.2
CH3Neu	0.000474	0.000341	0.20	0.04	1.71	1.35	1.4	1.0
H1Gal1	0.001735	0.000657	0.00	-0.41	-6.90	-2.04	5.0	1.9
H2Gal1	0.001027	0.000000	0.26	0.00	3.95	0	3.0	0.0
H3Gal1	0.000319	0.001068	-0.11	-0.51	-1.19	3.73	0.9	3.0
H4Gal1	0.001000	0.001260	0.22	0.21	3.88	4.94	2.9	3.6
H5Gal1	0.000465	0.000280	-0.46	-0.28	-0.27	0	1.3	0.8
H6Gal1a	0.001570	0.000000	1.57	0.00	0.00	0	4.5	0.0
H6Gal1b	0.000580	0.000650	-0.58	0.00	0.00	0	1.7	1.9
H1GlcNAc	0.001874	0.002033	0.89	1.03	-6.56	-6.97	5.4	5.8
H2GlcNAc	0.001637	0.001034	-0.18	-0.94	6.47	1.71	4.7	2.9
H3GlcNAc	0.000210	0.002057	-0.21	0.81	0.00	7.52	0.6	5.9
H4GlcNAc	0.000850	0.001195	-0.03	0.04	-3.38	4.75	2.5	3.4
H5GlcNAc	0.000000	0.000726	0.00	0.13	0.00	2.84	0.0	2.1
H6GlcNAc	0.001053	0.001586	-0.12	-0.49	4.16	12.4	3.0	4.5
H6GlcNAc	0.001393	0.001586	0.38	-0.49	5.33	12.4	4.0	4.5
CH3GlcNAc	0.000069	0.000401	0.07	0.21	0.00	1.35	0.2	1.1
H1Gal2	0.000722	0.000221	0.38	0.22	2.44	0	2.1	0.6
H2Gal2	0.001033	0.000620	0.77	0.62	2.74	0	3.0	1.8
H3Gal2	0.000413	0.000712	-0.13	-0.23	-1.56	-2.68	1.2	2.0
H4Gal2	0.000460	0.000400	-0.46	-0.40	0.00	0	1.3	1.1
H5Gal2	0.000220	0.000573	0.22	-0.19	0.00	-2.15	0.6	1.6
H6Gal2	0.000970	0.000650	0.61	0.23	-3.00	-2.42	2.8	1.9
H1Glcβ	0.000000	0.000306	0.00	-0.23	0.00	0.8	0.0	0.9
H1Glcα	0.000000	0.000000	0.00	0.00	0.00	0	0.0	0.0
H2Glcβ	0.000852	0.000184	0.25	0.18	3.24	0	2.5	0.5
H3Glcβ	0.000000	0.000000	0.00	0.00	0.00	0	0.0	0.0
H4Glcαβ	0.003614	0.003925	-0.24	-0.31	14.34	15.56	10.5	11.2
H5Glcβ	0.001018	0.000562	-0.16	0.34	4.00	-1.78	2.9	1.6
H6-Glcβ	0.001802	0.002026	0.08	-0.24	7.16	8	5.2	5.8
H6-Glcβ	0.001832	0.000724	-0.34	-0.53	7.16	1.96	5.3	2.1
NH-GlcNAc	nd	nd	8.28	8.0	nd	nd	nd	nd
NH-Neu5Ac	nd	nd	6.77	6.6	nd	nd	nd	nd

Table 2. Determination of temperature coefficients (ppb/K) and chemical shifts deviations (CSD) of LSTa and LSTc.

This behaviour suggested that the amide group from GlcNAc could be involved in an intra-molecular hydrogen bond. This result is supported by a study of the half-height line widths ($\Delta\nu_{½}$) extracted from each NH acetamide after a lorentzian deconvolution and plotted at different temperatures. NH from Neu5Ac N-acetyl group broadened before those of GlcNAc, suggesting that the NH of GlcNAc is involved in intra-molecular hydrogen bonds (Fig. 5).

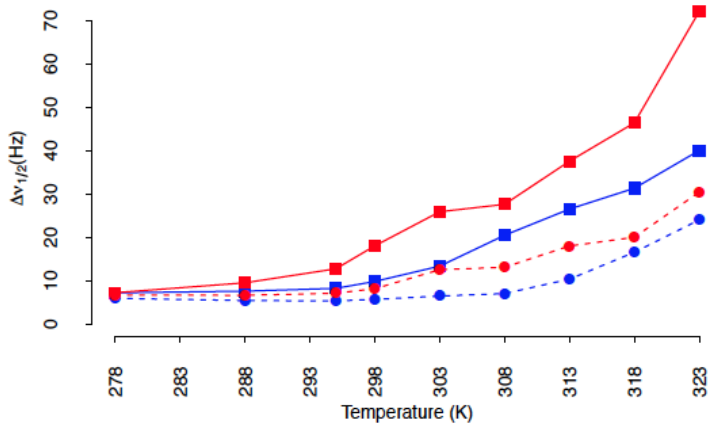


Fig. 5 Experimental curves of NH $\Delta\upsilon_{1/2}$ [Hz] plotted against temperature [K]. LSTc (square), LSTa (circle), Neu5Ac (red), GlcNAc (blu).

^{13}C NMR relaxation data. NMR relaxation studies have been utilized to interrogate the internal motion of the glycans, providing dynamic properties that can be qualitatively or quantitatively assessed [20]–[23]. T_1 and T_2 relaxation times and heteronuclear NOEs were measured at 14.0 T (600MHz) (Table 3). The numerical values of T_1 varied from 315 to 412 ms for LSTa and from 328 to 421 ms for LSTc. T_2 varied from 229 to 355 ms and from 232 to 387 ms for LSTa and LSTc, respectively.

LSTa	T_1	T_2	$R_{T1/T2}$	NOE	LSTc	T_1	T_2	$R_{T1/T2}$	NOE
C3-axNeu	159	135	1.18	1.50	C3-axNeu	157	138	1.14	1.75
C3-eqNeu	162	151	1.07	1.91	C3-eqNeu	140	155	0.90	1.79
C4-Neu	339	302	1.12	1.69	C4-Neu	331	299	1.11	1.79
C5-Neu	327	365	0.90	1.69	C5-Neu	341	218	1.56	1.76
C6-Neu	328	302	1.09	1.65	C6-Neu	335	235	1.43	1.77
C7-Neu	284	238	1.19	1.60	C7-Neu	315	270	1.17	1.59
C8-Neu	345	306	1.13	1.62	C8-Neu	361	329	1.10	1.86
C9-Neua	234.	200	1.17	1.79	C9-Neua	230	209	1.10	1.93
C9-Neub	230	190	1.21	1.81	C9-Neub	230	209	1.10	1.93
CH_3-NAcNeu	1770	22.2	79.7	1.26	CH_3-NAcNeu	2172	41	52.9	1.13
C1-Gal1	329	213	1.54	1.60	C1-Gal1	352	258	1.36	1.62
C2-Gal1	337	211	1.60	1.64	C2-Gal1	330	277	1.19	1.57
C3-Gal1	289	205	1.41	1.67	C3-Gal1	340	233	1.46	1.44
C4-Gal1	299	257	1.16	1.57	C4-Gal1	297	261	1.14	1.45
C5-Gal1	351	258	1.36	1.69	C5-Gal1	320	247	1.30	1.64
C6-Gal1a	220	203	1.08	1.76	C6-Gal1a	176	135	1.30	1.51
C6-Gal1b	220	203	1.08	1.76	C6-Gal1b	192	150	1.28	1.41
C1-GlcNAc	323	238	1.36	1.60	C1-GlcNAc	335	267	1.25	1.65
C2-GlcNAc	296	240	1.23	1.69	C2-GlcNAc	326	226	1.44	1.67
C3-GlcNAc	318	193	1.65	1.53	C3-GlcNAc	329	266	1.24	1.71
C4-GlcNAc	316	328	0.96	1.62	C4-GlcNAc	324	176	1.84	1.62
C5-GlcNAc	324	234	1.38	1.62	C5-GlcNAc	418	224	1.87	1.68
C6-GlcNAca	220	203	1.08	1.76	C6-GlcNAca	263	224	1.17	1.86
C6-GlcNAcb	220	203	1.08	1.76	C6-GlcNAcb	263	216	1.22	1.44
CH_3-GlcNAc	177	22.2	79.7	1.26	CH_3-GlcNAc	2190	23.2	94.4	1.93
C1-Gal2	333	269	1.24	1.76	C1-Gal2	374	153	2.44	1.80
C2-Gal2	362	314	1.15	1.83	C2-Gal2	359	310	1.16	1.74
C3-Gal2	334	214	1.56	1.67	C3-Gal2	335	249	1.35	1.78
C4-Gal2	354	379	0.93	1.64	C4-Gal2	318	342	0.93	1.62
C5-Gal2	348	264	1.32	1.88	C5-Gal2	365	297	1.23	1.81
C1-Glcα	319	329	0.97	1.60	C1-Glcα	340	306	1.11	1.54
C2-Glcα	442	334	1.32	1.98	C2-Glcα	455	502	0.91	1.83
C3-Glcα	433	389	1.11	1.97	C3-Glcα	395	385	1.03	2.00
C4-Glcαβ	385	275	1.40	1.86	C4-Glcαβ	416	249	1.67	1.86
C5-Glcα	429	351	1.22	1.82	C5-Glcα	414	485	0.85	1.90
C1-Glcβ	434	403	1.08	1.86	C1-Glcβ	470	372	1.26	1.86
C2-Glcβ	444	402	1.10	1.76	C2-Glcβ	458	411	1.11	1.93
C6-Glcβa	285	194	1.47	1.67	C6-Glcβa	191	191	1.00	1.79

Table 3. Spin.lattice and spin spin relaxation times (T_1 and T_2 values in ms) and heteronuclear NOEs of LSTa and LSTc measured at 600 MHz in acqueous solution at 295K.

In order to qualitatively observe the main differences of motional properties between LSTc and LSTa, the average values of T_1 and T_2 of each CH ring were used (Fig. 6).

The numerical values of T_1 for the different residues suggest an anisotropically tumbling molecule, since the Neu5Ac-Gal1-GlcNAc trisaccharide of both glycans gave rise to T_1 values ranging from 315 ms to 346 ms, whereas, values of the reducing Glc residue varied from 410 to 421 ms.

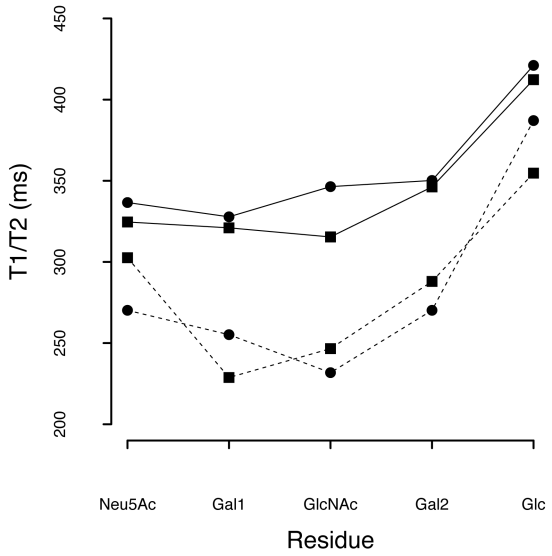


Fig. 6 Average ^{13}C relaxation times of CH carbons for each monosaccharide unit from LSTc (●) and LSTa (■). The solid line represents the average T_1 of each monosaccharide, while the dashed line represents the average T_2 of each monosaccharide.

This finding indicates a higher degree of motion for the reducing lactosyl disaccharide (Gal2-Glc) compared to the non-reducing end of the oligosaccharide. The heteronuclear NOEs corroborate the T_1 values, since larger

heteronuclear NOEs are observed for the terminal units, as reported in Tab3. Using the same approach, significantly lower T_2 values were detected for both LSTa and LSTc, T_2 values of the reducing Glc residue of both glycans (355 ms and 387 ms for LSTc and LSTa, respectively) confirmed the higher degree of motion of this residue compared to the internal Gal1 and GlcNAc residues showed lower T_2 values between 228 ms and 255 ms for both glycans. The main difference between LSTa and LSTc were the values observed for T_2 of the Neu5Ac residues, being 303 ms and 270 ms, respectively, such a difference, higher than experimental error (from 5 to 15ms), accounts for the significantly higher rigidity of the non-reducing end of LSTc. We want to underline how the higher mobility of the reducing end residues has little biologically relevance, since this part of the molecule *in vivo* is part of a long glycan chain. A relevant disparity between LSTa and LSTc is observed in the T_2 relaxation time of the non-reducing Neu5Ac residues, 303 and 270 ms, respectively. Such a difference, which is not related to chain length and is substantially higher than experimental error, accounts for the significantly higher rigidity of the non-reducing end of LSTc, as can be understood by the intra-residue hydrogen bond analysis section applied on MD simulation trajectories of LSTc and LSTa.

Inter- and intra- NOESY experimental distances. The inter- and intra-residue proton-proton distances were measured by NOESY enhancement. Intra residue distances from H1-H3 of Hex*p* units and H3-H5 of the Neu5Ac residues were measured between 2.0-2.6Å for all the residues (Table 4). These values were in agreement with the expected $^4C_1$ and $^2C_5$ ring chair conformations, supporting the observed $^3J_{H\text{-}H}$ coupling constants. Inter proton-proton distances were

obtained mainly from O-glycosidic bonds (Table 5). LSTa exhibited NOE contacts for $H3_{eq}$/$H3_{ax}$-Neu:H3-Gal1 providing distances of 4.7 and 2.8Å respectively, while, LSTc exhibited an inter residue NOE between H5-Neu5Ac and CH_3-GlcNAc corresponding to an estimated distance of 4.2 Å (Fig. 7 and 8).

LSTa	**Neu5Ac(*)**	**Gal1**	**GlcNAc**	**Gal2**	**Glc**
H1-H3	-	2.7/**2.4**	2.6/**2.0**	2.7/**2.2**	2.7
H3-H5	2.6/**2.6**	2.5	2.7	2.5	2.6
LSTc	**Neu5Ac(*)**	**Gal1**	**GlcNAc**	**Gal2**	**Glc**
H1-H3	-	2.7/**2.1**	2.6/**2.1**	2.7/**2.5**	2.7
H3-H5	2.6/**2.3**	2.54	2.6	2.5	2.7

Table 4 Selected experimental and predicted intra-residue proton-proton distances from LSTa and LSTc. Avarage value of distances H1-H3 and or H3—H5 for each residue calculated on 100 ns MD simulation. In bold face the corresponding experimental distances are reported as obtained by the NOESY extrapoleted at short mixing time. (*) For Neu5Ac the considered distances is the H3ax-H5.

This contact between GlcNAc and Neu5Ac residues (Fig.8), observed only in the NOESY spectra of LSTc, can be considered the experimental proof of a bent conformation that LSTc assumes compared to LSTa, which does not exhibit this signal at all. In fig. 7 there are the representation of the NOEs observed for LSTa (top) and LSTc (bottom) respectively.

In addition, we recorded also selective NOE experiments (Fig.9) of LSTc and LSTa at 900 MHz at different mixing times, and we compared them with the theoretical NOE signals generated by proposed models for LSTa and LTc using NOEPROM software (see also *NMR data and MD simulation comparison* section).

For same selected inter-glycosidic linkage, we reported the NOE intensity as a function of mixing time (Fig. 10).

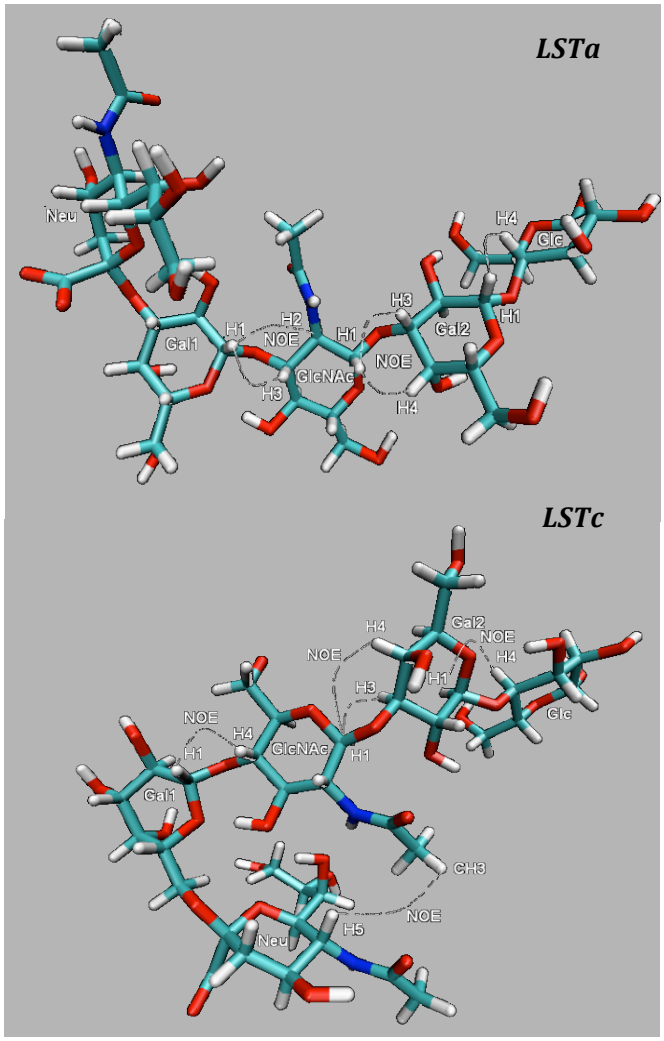


Fig. 7 3D model of LSTa (top) and LSTc (bottom) with the correspondents NOEs.

The theoretical values (square) obtained using NOEPROM are in agreement with the experimental data (circle), confirming moreover the validity of the molecular dynamic generated model (see also NMR data and MD simulation comparison).

	LSTc (Å)	LSTa (Å)
H3ax-Neu5Ac-H3-Gal1	-	2.76
H3eq- Neu5Ac-H3-Gal1	-	4.70
H5-Neu5Ac-CH_3GlcNAc	4.24	-
H1-Gal1-H3-GlcNAc	-	2.25
H1-Gal1-CH_3-GlcNAc	-	4.00
H1-GlcNAc-H3-Gal2	1.94	1.95
H1-GlcNAc-H4-Gal2	3.05	2.84
H1-Gal2-H4-Glc	1.96	2.20
H1-Gal2-H5-Glc	-	2.80

Table 5 - Selected inter-glycosidic proton-proton distances (Å) for LSTa and LSTc derived from NOE measurements.

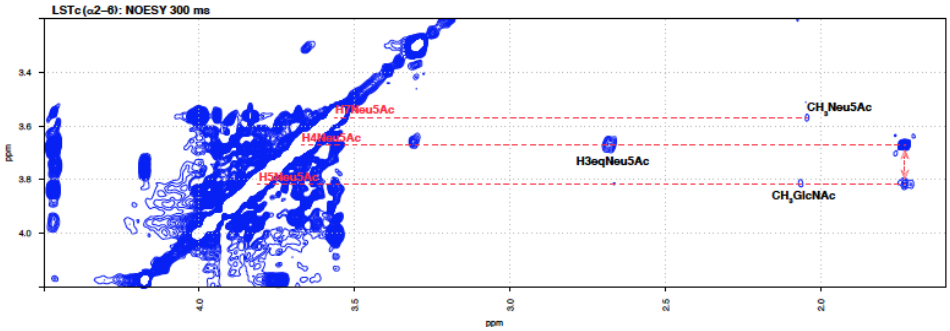


Fig. 8. 2D Partial NOESY (300 ms at 295K spectrum of LSTc showing the inter-residue NOE contact of H5Neu5Ac and CH3GlcNAc and the intra-residue NOE contacts of H4-, H5- and H7-Neu5Ac in the diagonal line.

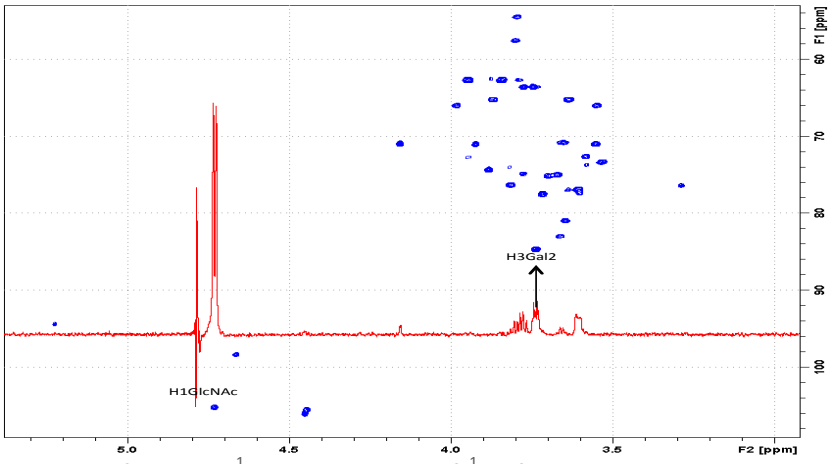


Fig. 9. Example of selective ^{1}H-noe experiment of ^{1}H of GlcNAc in which is visible the noe correlation with the H3 of Gal2.

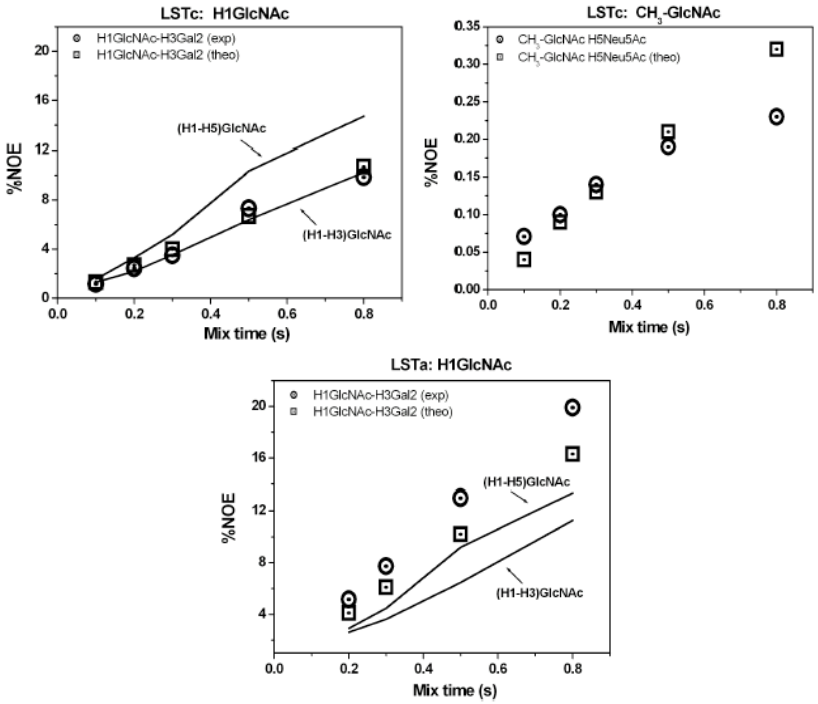


Fig. 10 % NOE of some selected residue for LSTa and LSTc. In circle the experimental values, and in square the theoretical values of the inter-glycosidic linkage. In continuos line the intra-residues.

The inter-glycosidic NOE between the CH_3 of the GlcNAc and the H5 of the Neu5Ac was further confirmed with 1D selective experiments.

Molecular dynamic simulations of LSTa and LSTc. To complement the NMR experiments, MD simulations were performed to compare the dynamics of the glycans in solution. The MD simulations were allowed to run for 100 ns, sampling conformers every 0.01 ns, therefore, producing a comprehensive dataset, with the simulation length considerably greater than the estimated correlation time (isotropic motion) of either glycan, of about 400 ps. After 70%

of the simulation time had elapsed several molecular parameters exhibited stationary behaviour without significant drift, these include the end-to-end distance of the sugars, which is defined as the distance between C5 of Neu5Ac (non-reducing) and C1 of Glc (reducing end), this property is one of the slowest evolving degrees of freedom for an oligomer.

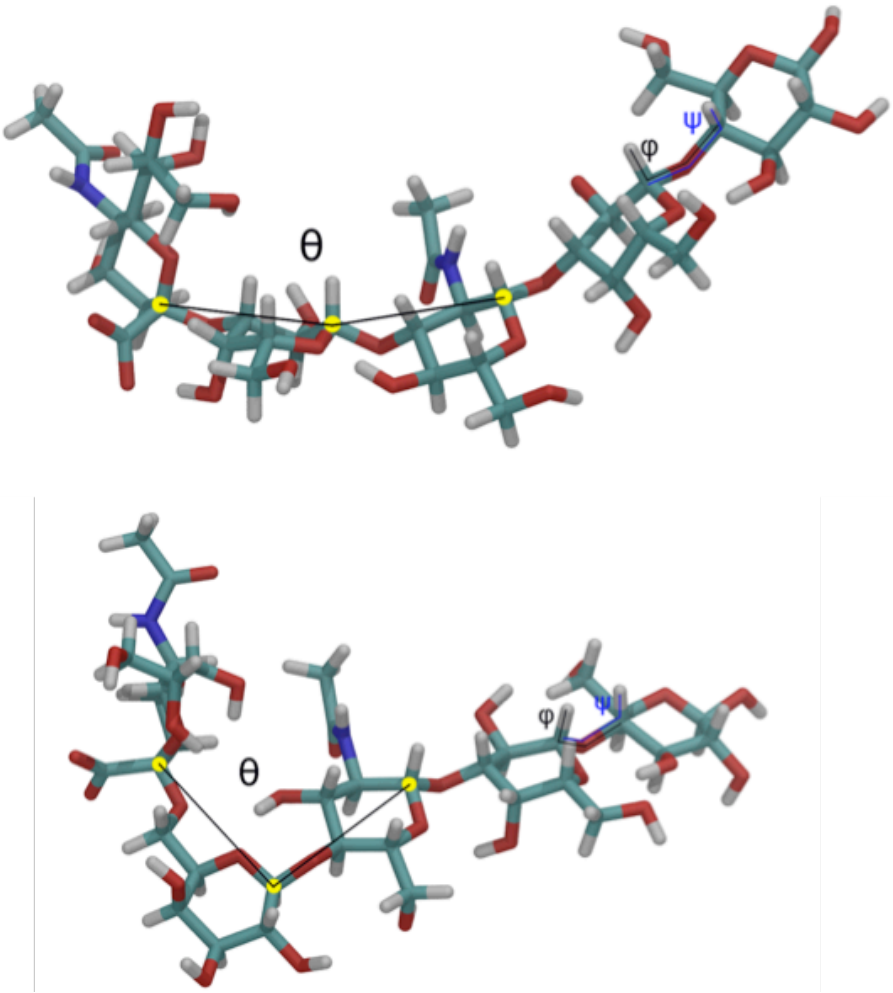


Fig. 11 3D model of LSTa (top) and LSTc (bottom) with the correspondent θ angles.

Stationary behaviour is also observed for the θ angle of the glycans (Fig. 11), which is characteristic for each glycan (Fig. 12 and 13) and the ω torsional angle of Gal1, which is defined the H5-C5-C6-O6 angle of Gal1 (Fig. 14).

Similar behavior was seen over the same timescale for the other four pairs of glycosidic torsional angles (ϕ_i/ψ_i), the angles that define the backbone conformation of the two oligomers. In this work, the torsional angle (ϕ and ψ) are defined by the following pairs of dihedral angles, show in fig. 11, from the non-reducing termini: ϕ_1 and ψ_1, ϕ_2 and ψ_2, ϕ_3 and ψ_3, and ϕ_4 and ψ_4. For LSTa, ϕ_1 is the C1-C2-O3-C3 angle and ψ_1 is the C2-O3-C3-H3 angle, while for LSTc, ϕ_1 is the C1-C2-O6-C6 angle and ψ_1 is the C2-O6-C6-C5 angle in agreement with Xu et al[18].

The successive pairs are defined as ϕ_1 and ψ_1 [H1-C1-O4'-C4' and C1-O4'-C4'-H4' (1-4 linkage) or H1-C1-O3'-C3' and C1-O3'-C3'-H3' (1-3 linkage)], as conventionally defined for sugars.

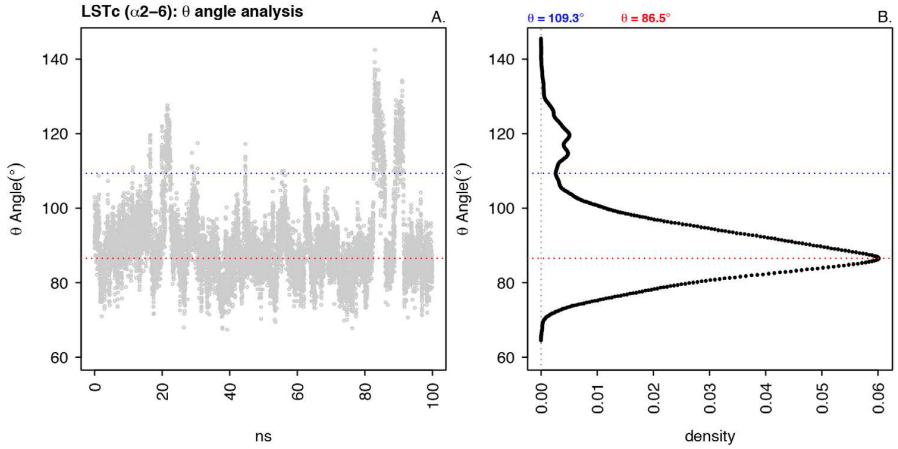


Fig. 12. LSTc θ angle. A. θ angle, B. density of θ angle.

Atoms labeled with a prime belong to the monosaccharide on the reducing side of the glycosidic linkage, while atoms without a prime are on the non-reducing side of the glycosidic linkage.

During the MD simulation the residues comprising LSTc and LSTa adopted the $^{2}C_{5}$ (Neu5Ac) and $^{4}C_{1}$ (Gal, Glc and GlcNAc) conformations, with only small deviations from the ideal chair of cyclohexane, the intra-residue proton-proton distances: H1-H3, H1-H5 and H3-H5 have values ranging between 2.5 - 2.7 Å for the five sugars, compared to the ideal value of 2.64 Å for cyclohexane. The same distances are estimated by NOE enhancement using the two-spin approximation (see Experimental Procedures), the values obtained ranged between 2.0 - 2.6 Å in good agreement considering the strong coupling effects (Table 5).

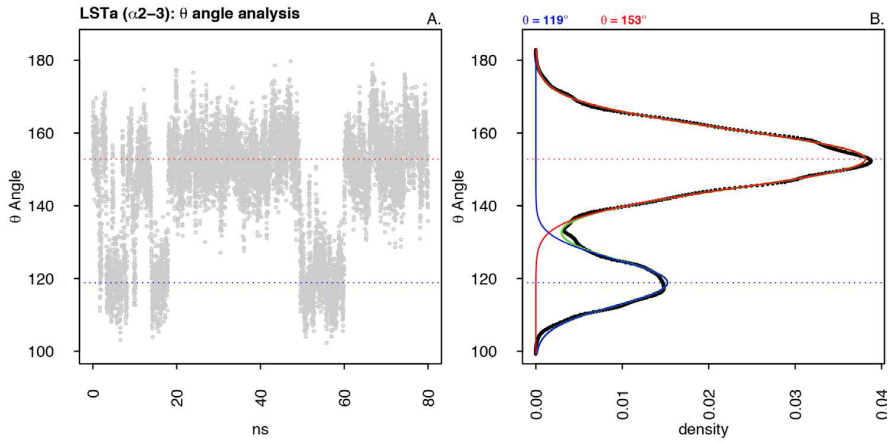


Fig. 13. LSTa θ angle. A. θ angle, B. density of θ angle.

The inter-sugar unit conformations were defined using four pairs of inter-residue glycosidic dihedral angles, running from the non-reducing to reducing-end of the glycans (ϕ_i/ψ_i).

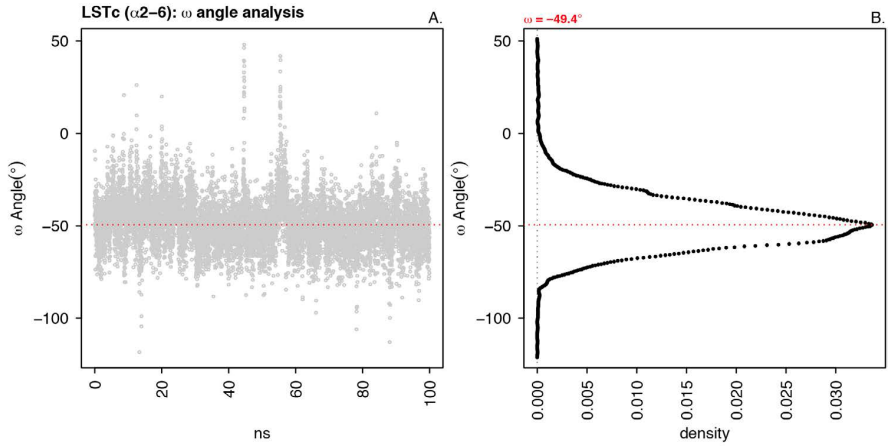


Fig. 14 LSTc ω angle. A. ω angle, B. density of ω angle.

The torsional angle values explored during the whole MD simulation are reported in Figure 15, the torsional angles are displayed in the range -120 to 240°. Visited torsional states are reported as cluster centroids with their relative populations in Table 6. The principal difference between LSTa and LSTc is located at the first glycosidic linkage (ϕ_1/ψ_1), which is in agreement with those obtained by molecular modelling [24], and is related to the different connectivity at the non-reducing end, LSTa - Neu5Ac-α(2→3)-Gal and LSTc - Neu5Ac-α(2→6)-Gal. For LSTa four ϕ_1/ψ_1 torsional states are observed, a principal state is located at -62°/-8° (89.8%), while the most significant member of the minor states is located at -83°/-55° (7.7%) (Figure 15A α and β, respectively). LSTc possess two ϕ_1/ψ_1 torsional states, the cluster with the greater population, 96.2%, is located at -57°/190° and the minor at 56°/174° (3.8%) (Figure 15E ξ and ο, respectively).

To fully describe the first glycosidic linkage of LSTc (Neu5Ac-α(2→6)-Gal) the ω torsional angle, needed to be considered.

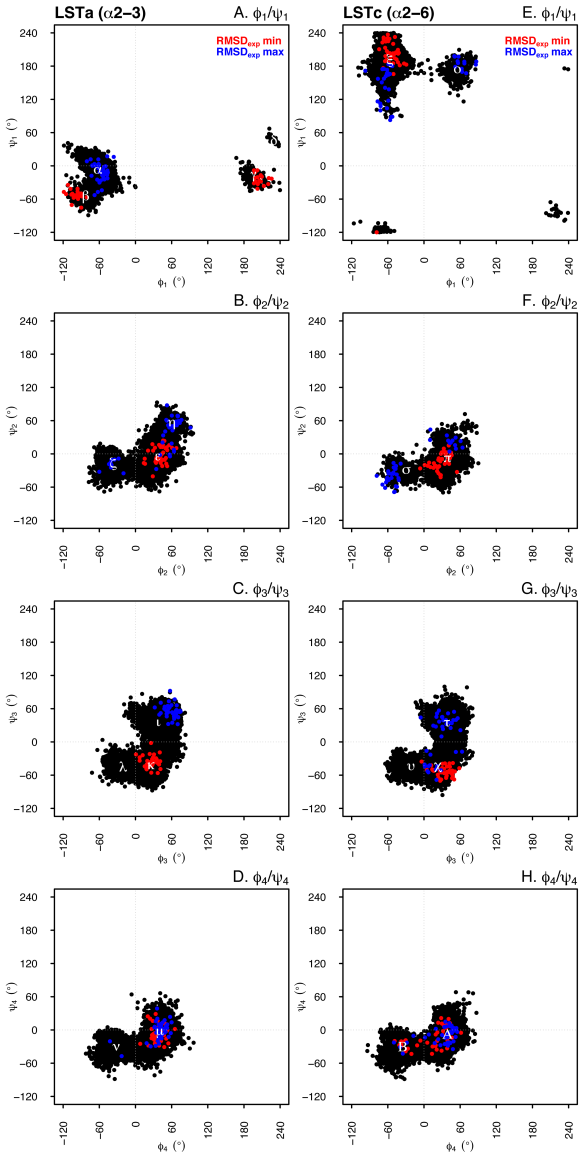


Fig. 15. Torsional states visited during the 100 ns simulation time for LSTa (left column) and LSTc (right column). For each glycan the dihedral angle pairs: ϕ_1/ψ_1 (**A**/**E**), ϕ_2/ψ_2 (**B**/**F**), ϕ_3/ψ_3 (**C**/**G**) and ϕ_4/ψ_4 (**D**/**G**) are shown. The red dots represent conformers that better fit the experimental NMR NOESY data ($RMSD_{exp}$min set of conformations), while blue dots correspond to the conformers that exhibit poor agreement with the experimental constraints ($RMSD_{exp}$ max). The positions and populations of the torsional states can be found in Table 6.

This degree of freedom is absent for LSTa. LSTc has a single state characterized by an average ω value of -50°, which is in agreement with the prediction of Xu et al. [18](Fig. 14). The torsional states determined for LSTa broadly agree with those described by Xu et al.[18], while the torsional angles determined for LSTc are less consistent with the data of the aforementioned work. Xu et al. [18]found additional LSTc torsional states located at, ϕ1/ψ1 ~180°/180°, ϕ2/ψ2 ~60°/180°, ϕ3/ψ3 ~ 30°/160° and ϕ4/ψ4 ~ 50°/180°. We predict a state in the first glycosidic linkage of LSTc located at 56°/174°, which is not observed by Xu et al. [18], but this state is seen in the conformational map obtained at GlycoMapsDB for Neu5Ac-α(2→6)-Gal[25]. This discrepancy is possibly related to the use of a different starting conformation for LSTc during the molecular dynamic simulation, which prevents an ensemble of conformations from being visited as a consequence of the potential energy walls restricting exploration of the conformational space.

To partially address this limitation, shorter MD simulations (20ns) were performed at higher temperatures, T=395 and 495K, using the same starting conformation at each temperature (Fig. 16 and 17). At these higher temperatures the missing LSTc torsional states in linkage ϕ_1/ψ_1 and ϕ_4/ψ_4 started to be populated (Fig. 16).

Interestingly, for both LSTc and LSTa the first glycosidic torsional state (ϕ_1/ψ_1) reproduces qualitatively in position and shape the torsional angle map of the glycosidic junction Neu5Ac-α(2→6/3)-Gal shown in GlycoMapsDB [25].

A parameter used to monitor the evolution of a MD simulation is the root mean square deviation (RMSD) of the atoms in a molecule (determined using the conformation at t=0s as reference), which illustrates the different

conformational states that the molecule assumes during the simulation, as compared to the starting state.

LSTa (α2-3) Torsional angles

linkage		ϕ (°)	ψ (°)	N°	%
1	α	**-62**	**-8**	8982	89.8
	β	**-83**	**-55**	770	7.7
	γ	**196**	**-19**	227	2.3
	δ	**230**	**47**	21	0.2
2	ε	**37**	**-5**	6908	69.1
	ζ	**-37**	**-18**	2402	24.0
	η	**60**	**54**	690	6.9
3	ι	**38**	**35**	5534	55.3
	κ	**26**	**-43**	3134	31.3
	λ	**-21**	**-44**	1332	13.3
4	μ	**40**	**-6**	7732	77.3
	ν	**-32**	**-31**	2268	22.7

LSTc (α2-6) Torsional angles

linkage		ϕ (°)	ψ (°)	N°	%
1	ξ	**-57**	**190**	9622	96.2
	ο	**56**	**174**	378	3.8
2	π	**40**	**-9**	9024	90.2
	σ	**-30**	**-30**	976	9.8
3	τ	**39**	**34**	4108	41.1
	χ	**23**	**-46**	4118	41.2
	υ	**-21**	**-45**	1774	17.7
4	A	**39**	**-7**	5840	58.4
	B	**-36**	**-30**	4160	41.6

Table 6. Torsional states of LSTa and LSTc as determined by molecular dynamics

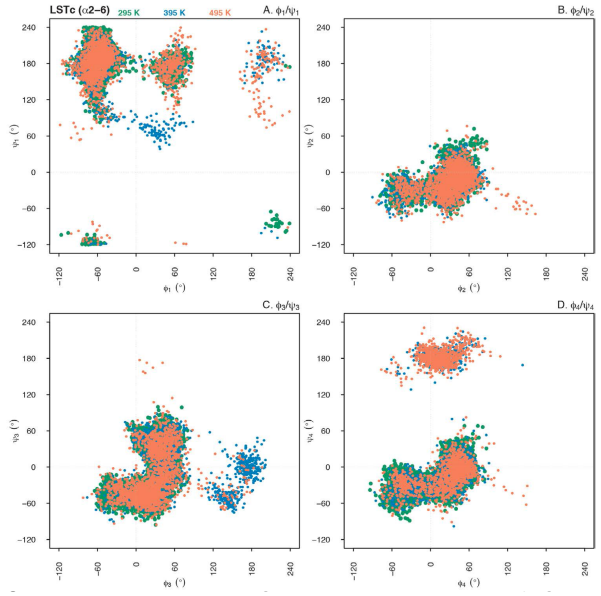


Fig. 16 Glycosidic torsional angles of LSTc at 295, 395 and 495 K. At 295 K the MD simulation length was 100 ns. While at 395 and 495 K it was 20 ns.

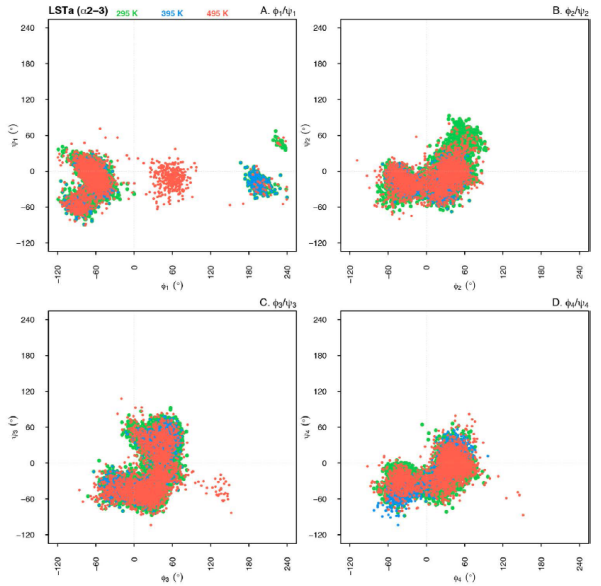


Fig. 17 Glycosidic torsional angles of LSTa at 295, 395 and 495 K. At 295 K the MD simulation length was 100 ns. While at 395 and 495 K it was 20 ns.

Unsurprisingly, the conformational states assumed by LSTc and LSTa are related to the torsional angle states of their glycosidic linkages.

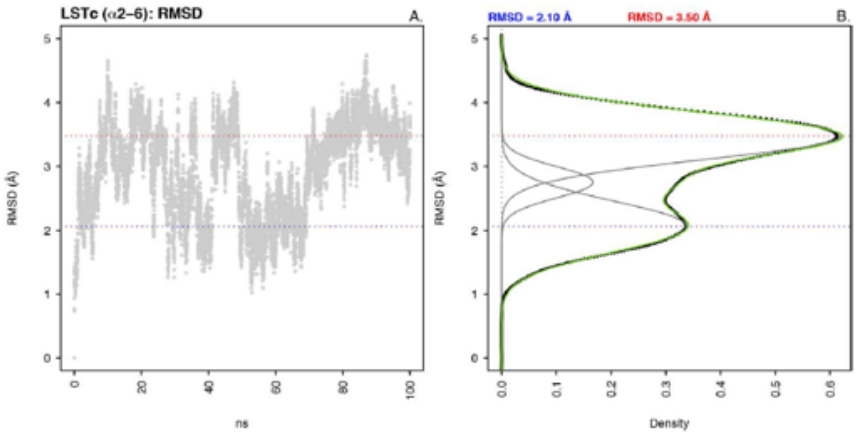


Fig. 18 RMSD analysis of LSTc over MD simulation, 100 ns, 295K. A. RMSD, B. density plot of the RMSD

During the simulation LSTc moves between two conformational states, RMSD states located at ~2.10 and ~3.50 Å (Fig. 18), while LSTa assumes a single broad distribution of related conformational states (Fig. 19) centred at ~2.8 Å.

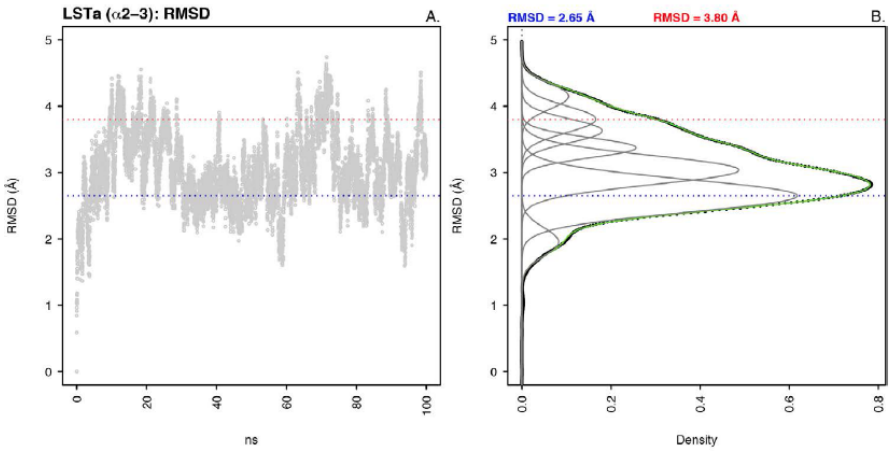


Fig. 19 RMSD analysis of LSTa over MD simulation, 100 ns, 295K. A. RMSD, B. density plot of the RMSD.

The conformers that populate the different RMSD states observed for LSTc, located at RMSD ~2.10 and ~3.50 Å, have stark differences in the third glycosidic linkage ϕ_3/ψ_3 (GlcNAc-β(1→3)-Gal-β) (Fig. 20). This correlation is also true for LSTa, a subset of conformers located at RMSD = 2.65 and 3.80 Å also exhibit significant differences in the third glycosidic linkage (Fig. 21). This result is interesting for the case of LSTc, as it suggests that the third glycosidic linkage is the most conformationally mobile junction and not the first, extended linkage Neu5Ac-α(2→6)-Gal, as might be assumed.

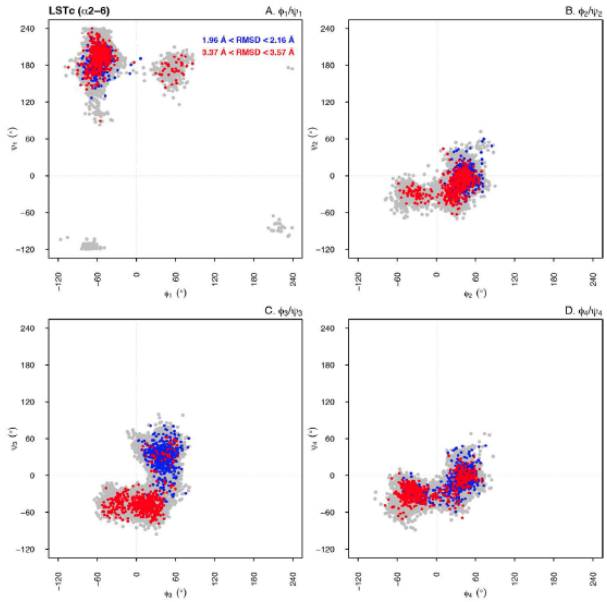


Fig. 20 Glycosidic torsional angles of LSTc with conformers from the different RMSD populations highlighted (blu 1.96 Å<RMSD<2.16Å and red 3.37Å<RMSD<3.57Å).

This could be related to the hydrogen bonds network that involve Neu5Ac with the rest of the molecule, the result being that the first linkage is less mobile compared to the third.

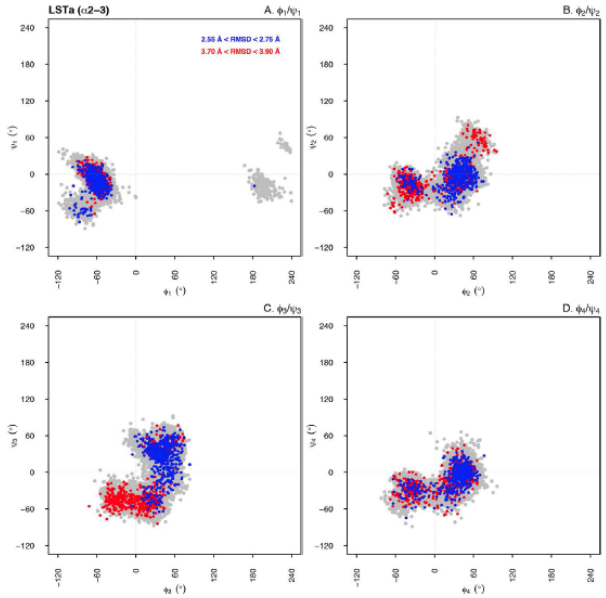


Fig. 21 Glycosidic torsional angles of LSTa with conformers from the different RMSW populations highlighted (blu 1.96 Å<RMSD<2.16Å and red 3.37Å<RMSD<3.57Å).

NMR data and MD simulation comparison. To compare the conformers derived by MD simulation and the experimental NMR data, a root mean squared deviation measure ($RMSD_{exp}$) was derived between 5 experimentally estimated inter-residue proton-proton distances and the corresponding MD predicted distances for each structure in the MD trajectories. The estimated inter-residue proton-proton distances for both LSTc and LSTa using the two spin model approximation, are reported for LSTc: H5-Neu5Ac:CH3-GlcNac, H1-GlcNAc:H3-Gal2, H1-GlcNAc:H4-Gal2, H1-Gal2:H4-Glc, H1-Gal1:H4-GlcNac, and for LSTa: H3ax-Neu5Ac:H3-Gal1, H3eq-Neu5Ac:H3-Gal1, H1-Gal1:H3-GlcNAc, H1-Gal1:CH3-GlcNac, H1-GlcNAc:H3-Gal2, H1-GlcNAc:H4-Gal2, H1-Gal2:H4-Glc, H1-Gal2:H5-Glc. In Eq. 1 $D_{\exp i}$ is the experimental i-inter proton-proton distance,

while $D_{MD\,i}$ is the corresponding distance for a single MD generated conformation. The number of the available inter-residue proton-proton distances is *n*.

$$RMSD_{exp} = \sqrt{n^{-1}\sum_{i=1}^{n}\left(D_{\exp i} - D_{MD\,i}\right)^2} \qquad (1)$$

The smaller the $RMSD_{exp}$ value the more closely the conformer matches the NMR data. The LSTc MD generated conformers have two distinct $RMSD_{exp}$ states, the first state located at 0.64 Å is very narrow and these conformers match the NMR data well, this population ends at 1.05 Å and thereafter there is a very broad declining population of conformers with increasing disparity between the NMR estimated inter-residue proton-proton distances and MD predicted distances. The conformers that populate these states have distinct glycosidic torsional states, with the conformers that more closely match the NMR data, the population with the smaller $RMSD_{exp}$ values ($RMSD_{exp}$ < 1.05 Å), belonging to the following glycosidic torsional angle states, ϕ_1/ψ_1 = -57°/190° and ϕ_2/ψ_2 = 39°/-9° (clusters ξ and π, respectively, Figure 15 E and F). LSTa also has two $RMSD_{exp}$ populations, located at 0.75 Å and 0.97 Å, again the conformers that comprise these different populations belong to differing torsional angle states. In this circumstance the differentiating linkage is closer to the reducing-end of the glycan, ϕ_3/ψ_3, the MD generated conformers that are in closer agreement to the NMR data are members of the torsional angle states located at ϕ_3/ψ_3 = -21°/-45° and 23°/-46° (clusters λ and κ, respectively, Figure 15C). It is interesting to note that on the whole the MD conformers of LSTa more closely match the NMR data, with the whole population falling

within the $RMSD_{exp}$ range 0 - 1.5 Å, whereas, the LSTc conformers have $RMSD_{exp}$ values spanning 0-5 Å. To further compare the experimental NMR data with the MD determined structures two subsets of conformers where chosen for each glycan, a subset containing the conformers with the 40 smallest $RMSD_{exp}$ values ($RMSD_{exp}$ min) [LSTc 0.24-0.33 Å, LSTa 0.12-0.43 Å] and a subset comprising the structures with the 40 largest $RMSD_{exp}$ values ($RMSD_{exp}$ max) [LSTa 1.26-1.41 Å, LSTc 3.83-4.40 Å]. The 2D-NOESY spectra were then simulated for these selected structures with the calculated intra- and inter-residue proton-proton NOE signals then being compared (Table 7). The intra-residue NOEs signals H1-H3, H1-H5 and H3-H5 are qualitatively reproduced for the LSTc residues Neu5Ac, Gal1, GlcNAc and Gal2, while for LSTa the signals are only reproduced for Gal1, GlcNAc and Gal2. Interestingly, the agreement between the simulated and experimental intra-residue 2D-NOEs signals is similar for both $RMSD_{exp}$ min and $RMSD_{exp}$ max conformer subsets, due to the $RMSD_{exp}$ values being determined using inter-residue distances and not intra-residue distances. In fact, the inter-residue distances H1Gal1-H4GlcNAc, H1GlcNAc-H3Gal2, H1Gal2-H4Glc and CH3GlcNAc-H5Neu5Ac compare well for the LSTc $RMSD_{exp}$ min conformer subset, while the $RMSD_{exp}$ max conformers only reproduce the H1Gal2-H4Glc distance. This trend is also true for LSTa, with the $RMSD_{exp}$ min set of structures reproducing the inter-residue distance H3Gal1-H3axNeu5Ac, H1Gal1-H3GlcNAc, H1GlcNAc-H3Gal2, and H1Gal2-H4Glc well, while the $RMSD_{exp}$ max conformers only replicate the experimentally determined distance H1Gal2-H4Glc adequately (Table 7). This illustrates the ability of the $RMSD_{exp}$ minimization procedure to select conformers from the MD simulation trajectory, providing a means of determining that quality of the MD simulation and allowing the structures that more closely match the experimental data to

be extracted. To get a broader comparison of the modeled and measured data, with MD conformers not being chosen by experimental constraints, the last 60 geometries of the MD simulation for LSTa and LSTc were used to generate NOE magnitudes (Table 7). For both glycans the agreement between the NOE magnitudes derived from MD and experimental data is qualitatively good, but the NOE magnitudes for the RMSDexp min MD conformer subset more closely match the experimental data than the NOE magnitudes derived from the MD structures, which were chosen without experimental constraints. It appears that the first glycosidic linkage, Neu5Ac-α(2→6/3)-Gal, for both glycan is the least well described by the MD simulation, while the remaining inter-glycosidic NOEs are better modeled in LSTa. This could be due to the isotropic motion approximation introduced by NOEPROM to generate the theoretical NOE magnitudes, with LSTc exhibiting stronger motional anisotropy compared to LSTa, which is observed in the NMR relaxation measurements.

The torsional angles for the $RMSD_{exp}$ min and $RMSD_{exp}$ max conformer subsets are shown in fig. 15, LSTa and LSTc, respectively. $RMSD_{exp}$ min structures are highlighted in red, while the $RMSD_{exp}$ max conformers are shown in blue.

Both of the conformer subsets for LSTc and LSTa show significant differences in their torsional angle states; ϕ_4/ψ_4 for both glycans is the only linkage where there is little divergence.

Interestingly, the LSTa $RMSD_{exp}$ min conformers have two ϕ_1/ψ_1 states, the linkage between Neu5Ac and Gal1, which are located at -100°/-53° and 207°/-24°, corresponding to the clusters β and γ, respectively. The conformers in the state centred at -100°/-53° are in better agreement with the NOE values for H3axNeu5Ac-H3Gal1 (Table 7). Sabesan et al. also observed a similar NOE, between H3axNeu5Ac-H3Gal1, for the di- and tri-saccharide structures[26].

		LSTc				LSTa	
CH_3GlcNAc		H5Neu5Ac		H3Gal1		H3axNeu5Ac	
Tm(s)	Exp.	$RMSD_{exp}$ min	$RMSD_{exp}$ max	Tm(s)	Exp.	$RMSD_{exp}$ min	$RMSD_{exp}$ max
0.1	0.07	0.04	0.00	0.20	1.26	2.10	0.10
0.2	0.10	0.09	0.00	0.30	1.12	3.20	0.20
0.3	0.14	0.13	0.00	0.50	1.55	5.40	0.30
0.5	0.19	0.23	0.00	0.80	2.21	8.90	0.50
0.7	0.23	0.32	0.00				
						$RMSD_{exp}$ min -99.5°/-52.9°	(ϕ_1/ψ_1) 206.6°/-23.8°
				0.20	1.26	0.41	4.10
				0.30	1.12	0.63	6.30
				0.50	1.55	1.10	10.6
				0.80	2.21	1.8	17.9
H1Gal1		H4GlcNAc		H1Gal1		H3GlcNAc	
0.15	3.80	3.60	0.80	0.20	4.44	5.20	1.90
0.2	5.20	4.80	1.00	0.30	6.09	7.80	2.90
0.3	6.80	7.30	1.5	0.50	11.07	13.10	4.80
				0.80	18.18	20.90	7.70
H1GlcNAc		H3Gal2		H1GlcNAc		H3Gal2	
0.1	1.16	1.40	1.90	0.20	5.15	7.20	0.70
0.2	2.40	2.70	3.70	0.30	7.73	10.80	1.00
0.3	3.50	4.10	5.60	0.50	12.98	18.00	1.70
0.5	7.30	6.80	9.30	0.80	19.91	28.80	2.80
0.8	9.90	11.00	14.90				
H1Gal2		H4Glc		H1Gal2		H4Glc	
0.15	2.40	2.80	2.50	0.20	3.80	5.70	5.20
0.2	2.60	3.70	3.30	0.30	5.90	8.60	7.80
0.3	3.40	5.50	5.00	0.50	10.30	14.30	13.00
				0.80	22.30	23.00	20.90

Table 7. Comparison between measured and simulated (NOEPROM) inter-residue proton-proton percentage NOE signals (%NOE). Percentage NOEs were simulated for the $RMSD_{exp}$ min and max groups for LSTc and LSTa, mixing times (Tm) span 0.1 to 0.8 s.

The glycosidic torsional angles of the conformers that more closely fit the NMR data are: LSTc - ϕ_1/ψ_1 = -55°/197°, ϕ_2/ψ_2 = 28°/-15°, ϕ_3/ψ_3 = 34°/-54°, ϕ_4/ψ_4 = 11°/-15° and LSTa - ϕ_1/ψ_1 = 207°/-24°, ϕ_2/ψ_2 = 41°/-1°, ϕ_3/ψ_3 = 27°/-34°, ϕ_4/ψ_4 = 38°/-5°.

To corroborate this result selected long range coupling constants, $^3J_{CH}$, across the glycosidic linkages (Jϕ = H1-C1-Ox-Cx and Jψ = C1-Ox-Cx-Hx) were measured using J-HMBC NMR experiment (Fig. 22), this allowed an indirect estimation of the glycosidic torsional angles for both glycans. Using the assignments obtained with HQSC spectra, it is possible to find in the 2D-J-HMBC

spectra the cross-peaks corresponding to the C-H correlation between different sugars. The 1D-extracted slice permit to calculate the $^{3}J_{CH}$ using a specific scaling factor (Fig. 22).

LSTa	$J^{exp}\phi_{(Hz)}$	$J^{calc}\phi_{(Hz)}$	$J^{exp}\psi_{(Hz)}$	$J^{calc}\psi_{(Hz)}$
Neu5Ac-Gal1 (ϕ_1/ψ_1)	-	-	4.4	2.2 (ϕ_1/ψ_1=-99°/-52°) 4.7 (ϕ_1/ψ_1=207°/-24°)
Gal1-GlcNAc(ϕ_2/ψ_2)	2.6	3.7	5.4	5.5
GlcNAc-Gal2 (ϕ_3/ψ_3)	4.2	4.8	4.5	3.8
Gal2-Glc (ϕ_4/ψ_4)	4.1	3.8	-	5.6
LSTc				
Neu5Ac-Gal1 (ϕ_1/ψ_1)	-	-	2.6	4.0
Gal1-GlcNAc(ϕ_2/ψ_2)	3.6	4.4	-	5.3
GlcNAc-Gal2 (ϕ_3/ψ_3)	4.1	3.9	4.4	2.2
Gal2-Glc (ϕ_4/ψ_4)	3.4	3.5 (ϕ_4/ψ_4=39°/-7°) 3.7(ϕ_4/ψ_4=-36°/-30°)	-	5.5 (ϕ_4/ψ_4=39°/-7°) 4.3 (ϕ_4/ψ_4=-36°/-30°)

Table 8. Experimental and calculated $^{3}J_{CH}$ interglycosidic coupling constant (Hz) for LSTa and LSTc (J_{ϕ}=H1-C1-O-Cx and J_{ψ}=C1-O-Cx-Hx).

The measured coupling constant (Hz) starting from the first torsion (ϕ_1/ψ_1) are for LSTa $J\psi_1$ = 4.4, $J\phi_2/J\psi_2$ = 2.6/5.4, $J\phi_3/J\psi_3$ = 4.2/4.5, $J\phi_4$ = 4.1 and for LSTc $J\psi_1$ = 2.6, $J\phi_2$ = 3.6, $J\phi_3/J\psi_3$ = 4.1/4.4, $J\phi_4$ = 3.4. Calculated pairs of couplings ($J\phi/J\psi$), determined using the $RMSD_{exp}$ min subset of conformers, in accordance to the Karplus-type equation [27], are reported in Table 8. The only measured long range coupling that does not agree with the $RMSD_{exp}$ min selected subset of conformers is ϕ_1/ψ_1 for LSTa, the measured coupling, ($J_{exp}\psi_1$ = 4.4 Hz), is in better agreement with the glycosidic torsional angles ϕ_1/ψ_1 = -99°/-53° and not ϕ_1/ψ_1 = 207°/-24°, which was determined by the NMR

selected MD conformers. J-HMBC spectra of LSTa and LSTc, overlaid with HSQC spectra, can be found in fig. 22.

Analyses of the shape and topology of LSTc and LSTa. The global shape of the two oligosaccharides is described by the topological θ angle (Fig. 11). Chandrasekaran et al. as previously described in chapter 1, proposed that the θ angle of LSTc and LSTa defines their conformation, with LSTa adopting a linear form, called "cone", which is defined as having a θ angle > 100°. While LSTc - Neu5Ac-α(2→6)-Gal-β(1→4)-, forms a bent conformation, called the "umbrella" form, defined as having a θ angle < 100° [17]. The LSTc conformers determined by MD have two θ angle populations, the much larger population is located at 87°, with the second smaller population starting above 109° (Fig. 12). These differences in θ angle also correspond to differences in the torsional angle states for the linkage between Gal1 and GlcNAc, the subset of conformers with a θ angle centered at 87° correlate with the torsional angle state ϕ_2/ψ_2 = 39°/-9°, while the structures with a θ angle greater than 88° populate torsional states belonging to $\phi_2/\psi 2$ = 39°/-9° and -30°/-30°.

This indicates that the second glycosidic linkage is important in the "opening and closing" of the umbrella conformation proposed for LSTc and not the first glycosidic linkage, which might be assumed because of the α2→6 linkage between Neu5Ac and Gal1. LSTa also has two θ angle populations, with the more populated state centered at 153° while the less populated state is located at 119° (Fig. 13). As for LSTc, the LSTa conformers belonging to the different θ angle populations correlate strongly with the torsional states of the second glycosidic linkage, ϕ_2/ψ_2.

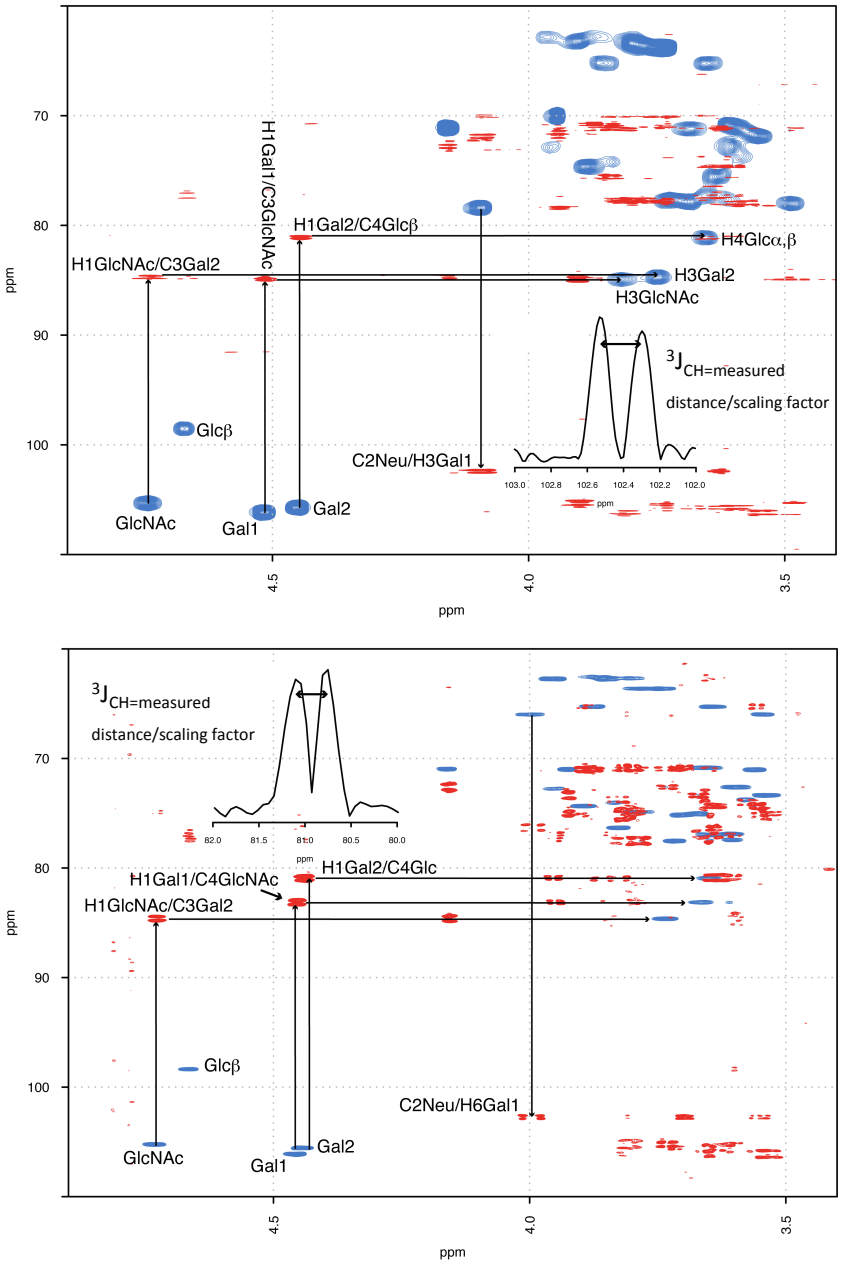


Fig.22 J-HMBC (red) and HSQC (blu) spectra of LSTa (top) and LSTc (bottom). The 1D-J-HMBC slice extracted for LSTa is the long range coupling for C2Neu/H3Gal1, while the 1D-J-HMBC slice shown for LSTc is the long range coupling costant for H1Gal2/C5Glc

The θ angle population centered at 153° correlates with the torsional state located at ϕ_2/ψ_2 = 37°/-5°, while the population located at 119° correlates with the torsional state ϕ_2/ψ_2 = -37°/18°. These results highlight that the conformational and subsequent dynamic properties of both LSTc and LSTa are correlated to the second glycosidic linkage. The LSTc conformers that more closely fit the NMR data ($RMSD_{exp}$ min structures) have an average θ angle of 87°, while the corresponding LSTa conformers have an average θ angle of 159°. These values, for the NMR selected MD generated conformers, agree with the results published by Xu et al. [18] and the hypothesis proposed by Chandrasekaran et al. [17].

Inter-residue hydrogen bond analysis. To further understand the consequences and driving-forces behind the different conformations assumed by LSTc and LSTa, the possible inter-residue hydrogen bonds were determined for the $RMSD_{exp}$ min and $RMSD_{exp}$ max conformer subsets. The analysis employed distance and angular restrictions; the distance between the donated hydrogen and the acceptor was ≤ 2.6 Å and the angle between X-H•••Y (X-donor and Y-acceptor) was greater than 110° tending towards 180°. We take into account the proposed hydrogen bonds for the $RMSD_{exp}$ min subset of conformers, the structures that more accurately represent the experimental data, the hydrogen bond analysis for the $RMSD_{exp}$ max conformers. For both glycans several possible hydrogen bonds were determined for the $RMSD_{exp}$ min conformers, 5 for LSTa and 11 for LSTc, illustrated in figure 23.

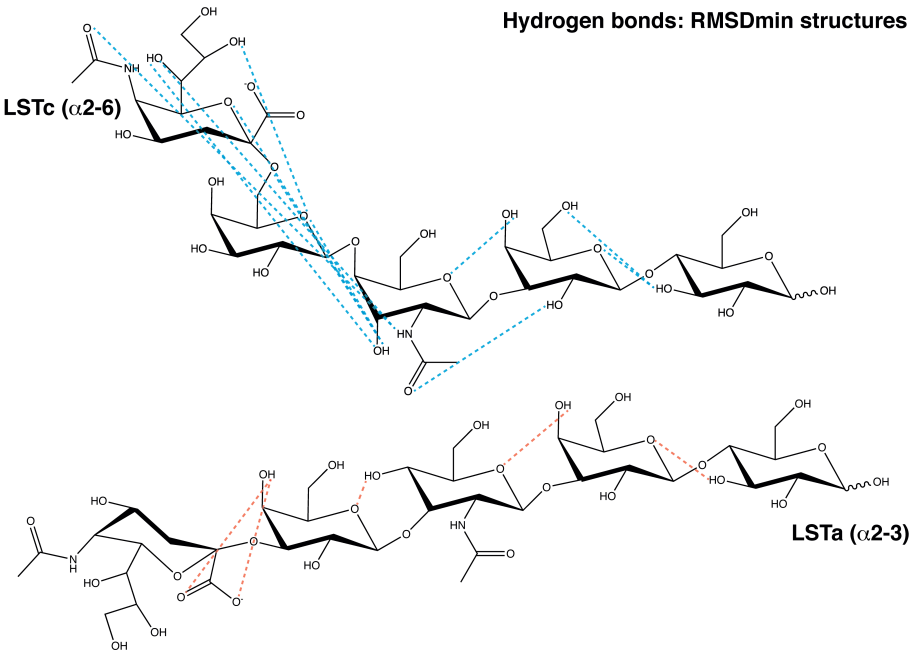


Fig. 23 Proposed hydrogen bond network for LSTc and LSTa $RMSD_{exp}$ min conformers. The hydrogen bond network at the non-reducing end of each glycan is strongly affected by the different linkage between Neu5Ac and Gal1 [LSTa, α(2→3) and LSTc, α(2→6)].

The hydrogen-bond networks present at the non-reducing end of either glycan are a direct consequence of their θ angles. The non-reducing end hydrogen bond network for LSTc, which has a bent form in solution (θ = 86° for the selected conformers), originates from GlcNAc, with two predicted bonds to Gal1 (H3O-GlcNAc:O5-Gal1 (100%), H3O-GlcNAc:O6-Gal1 (25.0%)) and 5 other spanning across Gal1 to Neu5Ac (O3-GlcNAc:H7O-Neu5Ac (55.0%), O5-Gal1:H8O-Neu5AC (20.0%), H2N-GlcNAc:O5N-Neu5Ac (15.0%), H3O-GlcNAc:O6-Neu5Ac (12.5%) and H2N-GlcNAc:O7-Neu (12.5%)) (Fig. 23). Owing to the bow-like backbone conformation of LSTa its hydrogen bonds can only be formed

between directly-adjacent residues: H4O-GlcNAc:O5-Gal1 (92.5%), H4O-Gal1:O1A/O1B (acid group)-Neu5Ac (52.5%), H4O-Gal2:GlcNAc (22.5%) and H3O-Glc:O5-Gal2 (80.0%) (Fig. 23).

Principal component analysis of the LSTa and LSTc MD trajectories. Principal component analysis (PCA) has been applied on the MD simulated trajectory to extract the independent mode of motion of both oligosaccharides, allowing an easier correlation with its structural properties. The convergence of the distinct ensembles of motion was observed by determining the rmsip (root-mean-square internal product) of the MD trajectory eigenvectors, comparing the eigenvectors generated by PCA of the whole MD tracjectory (100 ns) to the PCA of the trajectory at different time points (increasing steps of 1 ns). The rmsip is defined as :

$$rmsip = \sqrt{10^{-1} \sum_{i=1}^{10} \sum_{j=1}^{10} (\eta_i \, \nu_j)^2}, \quad (2)$$

where η_i is the ith eigenvector of trajectory η and ν_j is the *i*th eigenvector of trajectory ν.

The rmsip provides a measure of the similarity of the “essential modes” extracted; if the value is zero then they are dissimilar, while if the value is one the subspaces have converged. The rmsip for both LSTc and LSTa is greater than or equal to 0.9 for at least 50% of the trajectory, Amadei et al. [28] suggested that values greater than 0.7 provide evidence of good convergence.

The first 4 components generated by PCA for LSTa and LSTc explain 73.8% and 66.2% of the total variance, respectively.

Figures 24 and 25 provide an illustration of this mode of motion, “porcupine” representations of the first four eigenvector for LSTc and LSTa. The independent modes of motion described by component one for LSTc has the non-reducing end of the glycan close to stationary compared to the other residues, while the reducing-end (Gal2-Glc) rotates around their glycoside linkages, moving as a single unit. This is in agreement with the RMSD analysis of the LSTc MD trajectory.

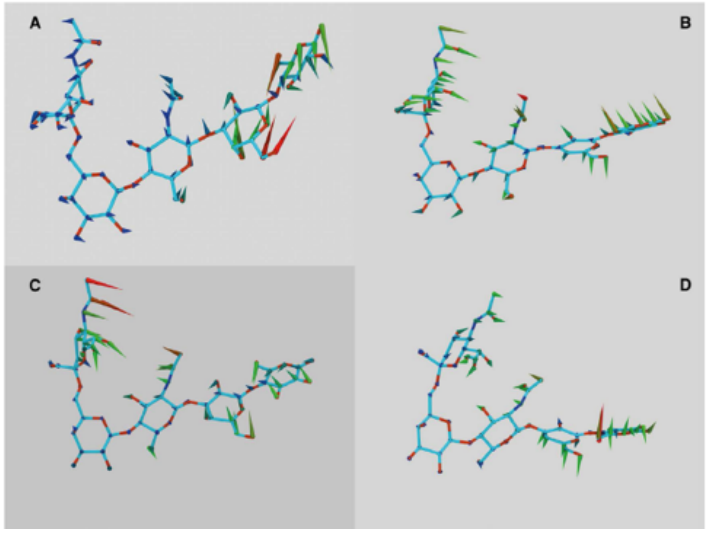


Fig. 24 Porcupine plot representations of the first four eigenvectors for LSTc. The first four distinct modes of motion revealed by PCA are illustrated in panels A-D. The direction, size and colour of the spines are proportional to the direction and the magnitude of the atom motion.

The three other distinct modes of motion extracted for LSTc show both the non-reducing and reducing ends of the molecule are in motion, with the reducing-end disaccharide (Gal2-Glc) moving as one unit and the sialyl group of Neu5Ac moving toward the N-acetyl group of GlcNAc (Fig. 24). The least mobile residue is Gal1, which exhibits the smallest variation in the 4 components examined. These results are consistent with a small θ angle. This dynamic

description agrees also with the results of the T_1 and T_2 relaxation measurements. The distinct modes of motion isolated for LSTa are very different from those of LSTc. LSTa has "bow-like" dynamics for the 4 components examined (Fig. 25), the whole molecule moving in unison, this is particularly true of the motion isolated for component 1.

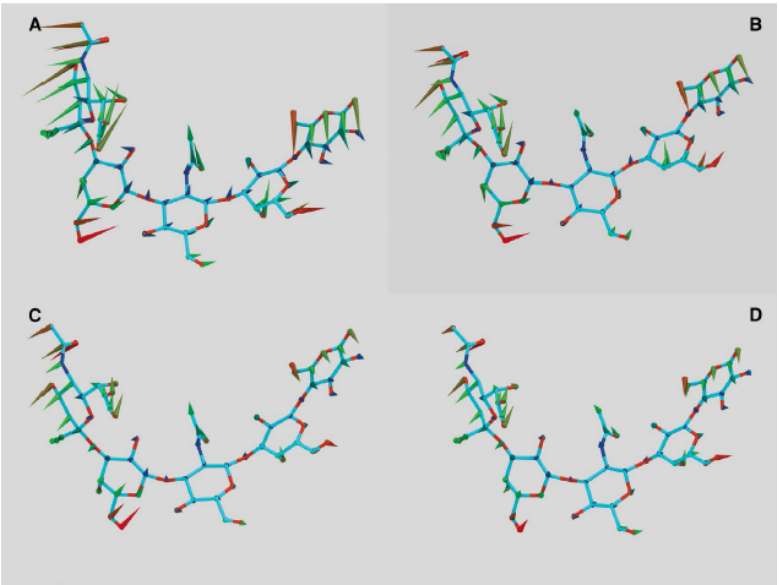


Fig. 25 Porcupine plot representations of the first four eigenvectors for LSTa. The first four distinct modes of motion revealed by PCA are illustrated in panels A-D. The direction, size and colour of the spines are proportional to the direction and the magnitude of the atom motion.

CONCLUSION

This study has identified the distinct conformational, topological and dynamic properties of the two pentasaccharides, used as a model for glycan receptors, LSTc and LSTa, of influenza A viruses free in solution. The study involved utilizing extensive NMR measurements – including the complete $^{1}H/^{13}C$ assignment for each glycan, molecular dynamic simulations and the numerical analysis of the MD generated trajectory to provide not only information concerning the conformation in solution (no bound state) but also, for the first time, details of its overall dynamic properties that could be related to the receptor binding ability.

Conformers from the MD simulations were selected using a root mean squared deviation ($RMSD_{exp}$) function, comparing NOE enhancement derived inter-glycosidic proton-proton distances and the corresponding predicted distance. Using this approach two subsets of conformers were selected, in good ($RMSD_{exp}$ min) or poor ($RMSD_{exp}$ max) agreement with the experimental data. In particular, the conformers which were in closer agreement with the experimental restraints, $RMSD_{exp}$ min, generated 2D-NOESY spectra, using the multi-spin approximation (NOEPROM), in good agreement with the values measured at 600 and 900MHz. These results were subsequently corroborated by experimental $^{n}J_{C,H}$ heteronuclear coupling constants, which were measured across the glycosidic linkages for LSTc and LSTa. LSTc exhibits a bent shape characterized by a topological angle, $\theta = 87°$, whereas LSTa has a more linear form and therefore a greater θ angle of 159°, a structural difference in agreement with the previous work of Chandrasekaran [17], Xu et al. [18] and concurs with the glycan ligand co-crystallized with HA [16].

As a consequence of these structural differences, the non-reducing end terminal Neu5Ac of LSTc could form an intricate hydrogen bond network with GlcNAc, which is not the case for its isomer LSTa, since hydrogen bond analysis revealed only interactions with directly adjacent residues.

In Figure 26 the structure of LSTc and LSTa extracted from the respective $RMSD_{exp}$ min sets are superimposed on the Neu5Ac and Gal1 residues in the co-crystallised glycan-HA complexes (right PDB: 2WRG 46 and left PDB: 1RVX 2).

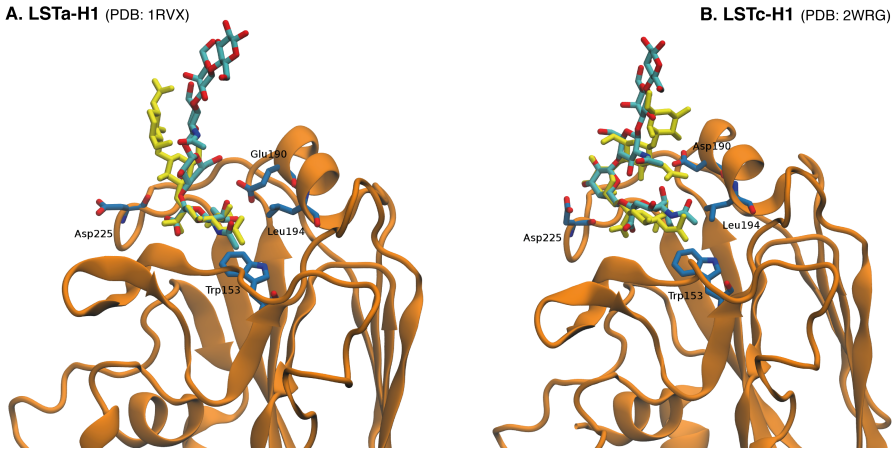


Fig. 26 Comparison of the MD generated LSTa (A) and LSTc (B) conformers (structures with the smallest $RMSD_{exp}$ value, $RMSD_{exp}$ min) and crystal structures of LSTa and LSTc bound to haemagglutinin from H1. The conformers generated though molecular dynamic simulations are cyan (C atoms) and red (H atoms), while the conformers determined by x-ray crystallography are yellow. The RMSD values between the molecular dynamic simulation generated structures and the co-crystallised glycans are 6.2 and 6.5 Å, LSTa and LSTc respectively.

RMSD values between the superimposed structures and the co-crystallised ligand, considering the three non-reducing end residues (i.e. Neu5Ac, Gal1 and GlcNAc) are 6.5 and 6.2 Å for LSTa and LSTc, respectively. These results illustrate how the NMR selected conformation for LSTc and LSTa in solution, i.e. determined without HA, match qualitatively the co-crystallised ligand bound to

HA, and also suggest that the respective sugar conformation does not significantly change on interacting with HA. This observation supports the hypothesis that HA recognizes the solution conformation of the glycan, instead to modify it during the interaction process. This study contributes to an improved understanding of the structural basis for glycan receptor specificity of HA in influenza.

EXPERIMENTAL PROCEDURES

NMR chemical shift mapping - NMR measurements were made using Bruker 500, 600 and 900 MHz AVANCE series NMR spectrometers (Bruker GmbH, Silberstreifen, Germany) with a high sensitivity 5 mm TCI cryoprobe. Human milk lactosyl sialyl-N-tetraoses namely LSTa (PROzyme, Hayward, CA, USA), (Neu5Ac-α(2→3)-Gal(Gal1)- β(1→3)-GlcNAc-β(1→3)-Gal(Gal2)- β(1→4)-Glc)) and LSTc (PROzyme, Hayward, CA, USA), (Neu5Ac-α(2→6)-Gal(Gal1)- β(1→4)-GlcNAc-β(1→3)-Gal(Gal2)- β(1→4)-Glc) were dissolved in D_2O (99.9%) or buffered solution (150 mM sodium chloride, 100 mM sodium phosphate, 10% D_2O, pH 7.2) according to the NMR experiment. Chemical shifts (^1H and ^{13}C) of the oligosaccharides were measured at 295 K in D_2O (99.9%) 2 mg / 0.2 ml using a 3 mm NMR tube. 2D NMR experiments were carried out using edited-HSQC, HSQC-TOCSY, HMBC, TOCSY and NOESY (using different mixing times between 100-1000 ms), recorded for quadrature detection in the indirect dimension and acquired using 16-48 scans per series of 1024x320 data points, with zero filling in F1 (4096) prior to Fourier transformation.

Estimation of intra and inter-proton distances - Experimental distances were obtained by selective 1D and 2D NOESY experiments recorded at mixing times between 100-800ms (295 K) and used to construct the kinetic NOEs. Selected distances were estimated based on the (two spin model approximation) time dependent NOEs measurements extrapolated to short mixing times, considering the initial build-up enhancement up to 300ms [29]. The extrapolated NOEs signals were converted to proton distances considering the distance Neu5Ac (H3ax-H3eq) as a reference (1.78 Å).

^{13}C NMR relaxation data - NMR relaxation times were obtained at 295K using a 14.0 T magnetic field. ^{13}C longitudinal relaxation (T_1) and transverse relaxation (T_2) were measured via 2D-double INEPT based inverse detection methods optimized to suppress the cross-correlation between the chemical shift anisotropy (CSA) and the dipolar relaxation. The 2D spectra were performed using 1024x192 data points, recording 24-32 scans at eight different relaxation delays arranged randomly to avoid systematic errors in every experiment. T_1 relaxation delays were namely 0, 0.1, 0.2, 0.3, 0.5, 0.8, 1.0, 1.4 s. T_2 transverse relaxation times were obtained by the CPMG pulse sequences differing in spin echo delays (22.4, 44.8, 67.2, 89.6, 134.4, 179.2, 224.0 and 313.6 ms). The volume of the cross-peaks were integrated using Topspin 2.1 and relaxation times were obtained by best non-linear and linear fitting of normalized data of the evolution of the peak volumes as a function of the inversion recovery delays. All the values and estimated errors (2-10%) were calculated and plotted in Origin 8.0 (OriginLab Corp., Northampton, MA, USA).

Determination of the $^{n}J_{(C,H)}$ heteronuclear coupling constants – Long range $J_{(C\text{-}H)}$ couplings were measured using the J-HMBC pulse sequence (295 K), with a

delta delay corresponding to a $^{n}J_{(X-H)}$ of 2 Hz (LSTc (0.7 mg) and LSTa (1 mg) were dissolved in 220 µL of D_2O). Data matrices of 1024 x 320 points covering 2003.2 x 9054.9 Hz were expanded 2048 x 1024 points using linear prediction and subsequently zero-filled to 8192 x 2048 points prior to echo/antiecho Fourier transformation. The traditional sequence was applied for LSTa, while the uniform sampling (NUS) variant was used for LSTc (Bruker Topspin 3.1).

Determination of temperature coefficients and chemical shift deviations - Temperature coefficients were recorded for 0.5 mg / 0.2 ml oligosaccharide solutions in buffered D_2O or in 90:10 H_2O:D_2O pH7.2. All temperature gradients were measured by differences observed in spectra recorded at 3 - 9 K intervals between 278 to 313 K (^{1}H). Temperature coefficients are expressed in units of ppb/K with a negative sign indicating an upfield shift upon warming, chemical shifts were measured using ^{1}H and HSQC experiments and chemical shift deviations were obtained using AMIX viewer 3.8 (Bruker GmbH, Silberstreifen, Germany) and global shifts deviations (CSD) determined using the equation:

$$\Delta\delta = \sqrt{(\Delta\delta_H)^2 + (\frac{\gamma^{13}C}{\gamma^{1}H}\Delta\delta_c)^2} \qquad (2),$$

Where $\Delta\delta_H$ and $\Delta\delta_c$ are the changes in proton and carbon chemical shifts and quantification was normalized by the gyromagnetic (γ) ratio.

Molecular dynamic simulations - The starting geometries of both oligosaccharides LSTa and LSTc were generated using the carbohydrate biomolecule builder (Woods Group (2005-2012) GLYCAM web), following the primary structure and the glycosidic constraints, both structures were minimized by the on-line energy minimiser (Woods Group (2005-2012)

GLYCAM web). The energy description of both oligosaccharides included the amber/glycam06 force field [30]. The atomic partial charges required by the oligosaccharides force field were calculated using the semi-empirical bond corrected AM1-BCC scheme[31]. The molecular dynamics (MD) simulations were conducted in explicit solvent using the TIP3P water model[32]. Oligosaccharides partial charge management and topology files were constructed with AmberTools 1.4. The MD trajectories were generated using the NAMD MD engine[33]. Non-bonded interactions i.e. electrostatic and van der Waals, are described using the cut-off approach set to 12.0 Å with a switching function beginning at 6.0 Å, the time step for the integration algorithm is 1 fs. The simulation cell including the oligosaccharide was built enveloping the solute molecule in a layer of water 20.0 Å wide in each direction; the result is a cubic cell of approximately 60 x 67 x 59 Å. The oligosaccharide models that were compared to the experimental data were generated by the following procedure; the simulation cell was first optimized applying 10000 minimization steps of the NAMD minimizer algorithm restraining each atom of the carbohydrate to the initial conformation by an harmonic restraint with a scaling constant of 500 Kcal/mol. The next step was a MD simulation reproducing the constant temperature and pressure statistical ensemble (NPT) applying a pressure on the cell walls of 1.01325 bar (atmospheric pressure). The simulation temperature was set at 295K and was controlled by a Langevin thermostat as implemented in NAMD. The Nosé-Hoover Langevin piston algorithm controlled the pressure applied on the cell walls and periodic boundary conditions were applied to the cell walls. The whole length of the simulation was 104 ns, where the first 4 ns were conducted restraining the carbohydrate atoms (solute) by an harmonic potential with a

scaling constant of 50 Kcal/mol to initially equilibrate the cell density without changing the solute conformation significantly, while the remaining simulation time was undertaken with the solute free to move. The cell densities calculated for LSTa and LSTc after equilibration (~ 4 ns) was in both cases 1.04 g/cm^3. The MD trajectory consisted of the last 100 ns, with a single geometry being sampled every 10 ps. During the whole dynamic, the conformation of the residues remained in $^2C_5$ for Neu5Ac and $^4C_1$ for the others, as set from the beginning for both LSTa and LSTc, and in agreement with the three bond proton-proton coupling constants ($^3J_{H\text{-}H}$).

The progression of the MD simulation were observed by determining the root mean squared displacement (RMSD) of the trajectories, performed in R using the rmsd function from the Bio3D library [34], with the starting conformation as the reference. The RMSD distances were determined using the following equation (2):

$$rmsd(t) = \sqrt{n^{-1} \sum_{i=1}^{n} \left[D_i(t) - D_i^{ref} \right]^2} \qquad (3)$$

Where D_i^{ref} is the position of the ith atom of the reference structure ($t = 0$), $D_i(t)$ is the position of atom i at time i and n is the number of atoms.

2D-NOESY spectra simulation (Multi-spin model relaxation) - The 2D-NOESY spectra was simulated using two different sets of 40 conformations each from the MD trajectory: 40 with the smallest RMSD ($RMSD_{exp}$ min) and highest ($RMSD_{exp}$ max) respectively. The 2D-NOESY simulation was undertaken assuming an isotropic motion with a single correlation time τ_c estimated at 400ps. The τ_c estimation was done on the smallest RMSD set of conformations

($RMSD_{exp}$ min) and simulating the intra-residue H1-H3 and H1-H5 2D-NOEs signals for the selected residue: GlcNAc for LSTc and Gal1, GlcNAc and Gal2 for LSTa. The τ_c parameter was adjusted by trial and error until the NOE kinetic enhancement from 0 until 500 ms mixing time, reproduced qualitatively the corresponding experimental values. The software for 2D-NOESY simulation was NOEPROM.

Principal component analysis of the MD trajectory - The MD trajectories for LSTc/LSTa were exported as XYZ coordinates via VMD [35], allowing the data to be imported directly in to R [36]. Principal component analysis (PCA) was performed using the bio3d library, the pca.xyz function [34]. The data were referred to the first conformer of the MD trajectory, removing translational and rotational effects from the analysis, and then mean centered before PCA was performed. The calculated eigenvalues provide a weighting for importance of the eigenvectors [illustrated as porcupine plots] and the component scores (score plot) illustrate the relationship between the determined eigenvector (independent molecular motion) and the conformers that constitute the MD trajectory. Porcupine plots were produced by constructing a trajectory along the desired principal component, imposing the molecular motion extracted by that component on to the average geometry of the conformers [34]. This was then visualized using VMD and the porcupine plot created using a VMD script[35].

REFERENCES

[1] G. L. Sassaki, S. Elli, T. R. Rudd, E. Macchi, E. A. Yates, A. Naggi, Z. Shriver, R. Raman, R. Sasisekharan, G. Torri, and M. Guerrini, "Human (alfa2->6) and avian (alfa2->3) sialylated receptors of influenza A virus show distinct conformations and dynamics in solution," *Biochemistry*, vol. 52, no. 41, pp. 7217–7230, 2013.

[2] M. B. Eisen, S. Sabesan, J. J. Skehel, and D. C. Wiley, "Binding of the influenza A virus to cell-surface receptors: structures of five hemagglutinin-sialyloligosaccharide complexes determined by X-ray crystallography.," *Virology*, vol. 232, no. 1, pp. 19–31, May 1997.

[3] S. J. Gamblin, L. F. Haire, R. J. Russell, and D. J. Stevens, "The Structure and Receptor Binding Properties of the 1918," vol. 303, no. March, pp. 1838–1842, 2004.

[4] Y. Ha, D. J. Stevens, J. J. Skehel, and D. C. Wiley, "X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 20, pp. 11181–6, Sep. 2001.

[5] J. Stevens, O. Blixt, T. M. Tumpey, J. K. Taubenberger, J. C. Paulson, and I. a Wilson, "Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus.," *Science*, vol. 312, no. 5772, pp. 404–10, Apr. 2006.

[6] W. D. C. Weis W., Brown J. H., Cusack S., Paulson J.C., Skehel J.J., "Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid," *Nature*, vol. 333, no. 2, pp. 426–431, 1988.

[7] A. Ibricevic, A. Pekosz, M. J. Walter, C. Newby, J. T. Battaile, E. G. Brown, M. J. Holtzman, and S. L. Brody, "Influenza virus receptor specificity and cell tropism in mouse and human airway epithelial cells.," *J. Virol.*, vol. 80, no. 15, pp. 7469–80, Aug. 2006.

[8] K. Shinya, M. Ebina, S. Yamada, M. Ono, N. Kasai, and Y. Kawaoka, "Avian flu: influenza virus receptors in the human airway.," *Nature*, vol. 440, no. 7083, pp. 435–6, Mar. 2006.

[9] D. van Riel, V. J. Munster, E. de Wit, G. F. Rimmelzwaan, R. a M. Fouchier, A. D. M. E. Osterhaus, and T. Kuiken, "Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals.," *Am. J. Pathol.*, vol. 171, no. 4, pp. 1215–23, Oct. 2007.

[10] A. Jayaraman, C. Pappas, R. Raman, J. a Belser, K. Viswanathan, Z. Shriver, T. M. Tumpey, and R. Sasisekharan, "A single base-pair change in 2009 H1N1 hemagglutinin increases human receptor affinity and leads to efficient airborne viral transmission in ferrets.," *PLoS One*, vol. 6, no. 3, p. e17616, Jan. 2011.

[11] K. Kumari, S. Gulati, D. F. Smith, U. Gulati, R. D. Cummings, and G. M. Air, “Receptor binding specificity of recent human H3N2 influenza viruses.,” *Virol. J.*, vol. 4, p. 42, Jan. 2007.

[12] M. N. Matrosovich, T. Y. Matrosovich, T. Gray, N. a Roberts, and H.-D. Klenk, “Human and avian influenza viruses target different cell types in cultures of human airway epithelium.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 13, pp. 4620–4, Mar. 2004.

[13] W. D. C. Rogers G.N., Paulson J.C., Daniels R.S., Skehel J.J., Wilson J.A., “Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity,” *Nature*, vol. 173, pp. 76–78, 1983.

[14] R. J. Russell, D. J. Stevens, L. F. Haire, S. J. Gamblin, and J. J. Skehel, “Avian and human receptor binding by hemagglutinins of influenza A viruses.,” *Glycoconj. J.*, vol. 23, no. 1–2, pp. 85–92, Feb. 2006.

[15] Z. Shriver, R. Raman, K. Viswanathan, and R. Sasisekharan, “Context-specific target definition in influenza a virus hemagglutinin-glycan receptor interactions.,” *Chem. Biol.*, vol. 16, no. 8, pp. 803–14, Aug. 2009.

[16] C. A. Bewley, “Illuminating the switch in influenza viruses,” *Nat. Biotechnol.*, vol. 26, pp. 60–62, Dec. 2008.

[17] A. Chandrasekaran, A. Srinivasan, R. Raman, K. Viswanathan, S. Raguram, T. M. Tumpey, V. Sasisekharan, and R. Sasisekharan, “Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin.,” *Nat. Biotechnol.*, vol. 26, no. 1, pp. 107–13, Jan. 2008.

[18] D. Xu, E. I. Newhouse, R. E. Amaro, H. C. Pao, L. S. Cheng, P. R. L. Markwick, J. A. McCammon, W. W. Li, and P. W. Arzberger, “Distinct glycan topology for avian and human sialopentasaccharide receptor analogues upon binding different hemagglutinins: a molecular dynamics perspective.,” *J. Mol. Biol.*, vol. 387, no. 2, pp. 465–91, Mar. 2009.

[19] C. J. Turner, P. J. Connolly, and a S. Stern, “Artifacts in sensitivity-enhanced HSQC.,” *J. Magn. Reson.*, vol. 137, no. 1, pp. 281–4, Mar. 1999.

[20] M. Hricovlni and G. Torri, “Dynamics in aqueous solutions of the pentasaccharide corresponding to the binding site of heparin for antithrombin III studied by NMR relaxation measurements 1 , 2,” *Carbohydr. Res.*, vol. 268, pp. 159–175, 1995.

[21] A. Poveda, M. Martin-pastor, M. Bernabe, J. A. Leal, and J. Jimenez-barbero, “Solution conformation and dynamics of a fungal cell wall polysaccharide isolated from Microsporum gypseum,” *Glycoconj. J.*, vol. 15, pp. 309–321, 1998.

[22] Eli, L. Poppe, and H. Van Halbeek, “The Rigidity of Sucrose: Just an Illusion?,” *J. Am. Chem. Soc.*, vol. 114, pp. 1092–1094, 1992.

[23] L. Poppe, H. Van Halbeek, D. Acquotti, and S. Sonnino, “Evolution of the receptor binding properties of the influenza A(H3N2) hemagglutinin,” *Biophys. J.*, vol. 66, no. May, pp. 1642–1652, 1994.

[24] M. Raab and I. Tvaroška, “The binding properties of the H5N1 influenza virus neuraminidase as inferred from molecular modeling.,” *J. Mol. Model.*, vol. 17, no. 6, pp. 1445–56, Jun. 2011.

[25] M. Frank, T. Lütteke, and C.-W. von der Lieth, “GlycoMapsDB: a database of the accessible conformational space of glycosidic linkages.,” *Nucleic Acids Res.*, vol. 35, no. Database issue, pp. 287–90, Jan. 2007.

[26] Sabesan S., K. Bock, and P. James, “Conformational analysis of sialyloligosaccharides *,” *Carbohydr. Res.*, vol. 218, no. 5900, pp. 27–54, 1991.

[27] E. P. Igor Tvaroska, Milos Hricovini, “An attempt to derive a new karplus type equation of vicinal proton carbon coupling constant for C-O-C-H segments of bonded atoms,” *Carbohydr. Res.*, vol. 189, pp. 359–362, 1989.

[28] A. Amadei, A. B. Linssen, and H. J. Berendsen, “Essential dynamics of proteins,” *Proteins*, vol. 17, no. 4, pp. 412–425, 1993.

[29] D. Neuhaus and M. P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VHC Publis. 1989.

[30] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. Gonza, C. R. Daniels, B. L. Foley, and R. J. Woods, “GLYCAM06 : A Generalizable Biomolecular Force Field .,” *J. Comput. Chem.*, vol. 29, pp. 1623–1641, 2007.

[31] A. Jakalian, D. B. Jack, and C. I. Bayly, “Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation.,” *J. Comput. Chem.*, vol. 23, no. 16, pp. 1623–41, Dec. 2002.

[32] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water,” *J. Chem. Phys.*, vol. 79, no. 2, p. 926, 1983.

[33] J. C. Phillips, E. Villa, and K. , Christophe Chipot, Robert D. Skeel, Laxmikant Kalé, and Schulten, “Scalable Molecular Dynamics with NAMD.pdf,” *J. Comput. Chem.*, vol. 26, no. 16, pp. 1781–1802, 2005.

[34] B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, and L. S. D. Caves, “Bio3d: an R package for the comparative analysis of protein structures.,” *Bioinformatics*, vol. 22, no. 21, pp. 2695–6, Nov. 2006.

[35] A. D. and K. S. William Humphrey, “VMD:Visual Molecular Dynamics,” *J. Mol. Graph.*, vol. 14, pp. 33–38, 1996.

[36] R. development core Team, “R: A language abd enviroment for statistical computing,” vol. 2. 2012.

Insights into Human Glycan Receptor conformation of 1918 Pandemic Haemagglutinin-Glycan Complexes derived from NMR and MD studies[1]

INTRODUCTION

Influenza A viruses are among the most rapidly evolving pathogens with potential for new strains to adapt to human hosts and result in a pandemic outbreak, with significant economic and public health impact [2][3]. The emergence of novel influenza strains such as 2009 H1N1 and 2010 H3N2 through genetic reassortment [4]–[6] pose a constant threat in terms of the evolution of various subtypes, including H5N1, H7N2, H7N7, H7N9, and H9N2 to generate a pandemic strain. The H5 and H7 subtypes, among others, are categorized as avian-adapted since they primarily circulate within birds, but, through contact with infected animals, they can jump species and infect humans. Some of these avian-adapted viruses, including H5N1 and H7N9, upon infection, can replicate efficiently in various human organs, and lead to severe infection and mortality [7]–[12]. However, these avian-adapted subtypes are not capable of efficient human-to-human aerosol transmission [13][14] a characteristic feature of subtypes such as H1N1, H2N2, and H3N2, which are considered human-adapted.

A key factor governing human adaptation of the influenza A virus is the binding specificity of viral surface glycoprotein haemagglutinin (HA) to sialylated glycan receptors on the host cell surface (glycans terminated by α-D-N-acetyl

neuraminic acid; Neu5Ac). A canonical definition of this binding specificity based on the terminal sialic acid linkage has been used in the field in recent decades. HA from avian-adapted subtypes such as H5, H7 and H9 is known to bind specifically to glycans terminated by α(2→3) linked sialic acid (α(2→3) glycans or avian receptors) [15][16]. Meanwhile, HA from human-adapted subtypes such as H1, H2 and H3 is known to bind specifically to glycans terminated by α(2→6) linked sialic acid [17][18]. This definition based on sialic acid linkage alone, while useful to characterize many influenza strains, has misclassified some notable virus strains in terms of their ability to effect efficient human-to-human respiratory droplet transmission [19][20][21], in particular H5N1 and H7N9.

Glycan receptors have been defined according to their overall conformation, via a parameter called the θ angle, which defines the form of the non-reducing end of the glycan receptors (Fig. 1) [19]. Specifically, in the case of influenza, we previously demonstrated that the apical surface of human upper respiratory epithelia, which is a primary target for human-adapted viruses, predominantly expresses glycan receptors with long oligosaccharide branches terminated by α(2→6)-linked sialic acid (α(2→6) glycan or human receptors) [19][22]. Based on analyses of glycan conformation and topology in HA-glycan X-ray co-crystal structures, we noted that the α(2→6) glycans adopted an 'umbrella-like' topology, θ angle < 100°, in the receptor-binding site (RBS) of H1 and H2 HAs, while α(2→6) and α(2→3) glycans adopted 'cone-like' topologies, θ angle > 100°, in the RBS of H5 HA[19]. This topology-based definition, in addition to the specific sialic acid linkage (i.e., α(2→3) vs. α(2→6)) distinguished HA from human-adapted subtypes binding to human receptors from the HA of avian-adapted viruses binding to avian and human receptors[23].

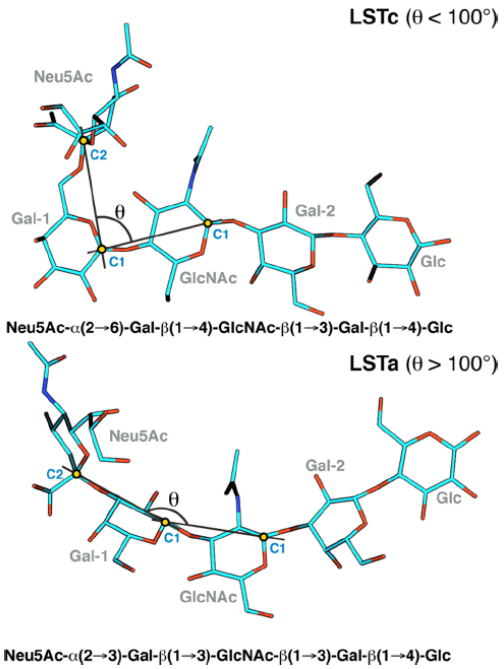


Fig. 1. Schematic of LSTc (top) and LSTa (bottom). The θ angle parameter is defined by the angle across anomeric carbons of the three successive residues starting from the non-reducing end (Neu5Ac, Gal1 and GlcNAc).

The relationship between glycan receptor specificity of HA and aerosol transmissibility of the virus in ferrets (a well-established animal model for respiratory droplet transmission) was first demonstrated using the prototypic 1918 H1N1 pandemic HA (A/South Carolina/1/18 or SC18) [24]. While SC18 showed efficient aerosol transmission in ferrets, a single amino acid mutation in HA (Asp225→Gly; numbering based on H3 HA) resulted in a virus, NY18, that exhibited inefficient transmission, and a second further mutation, Asp190→Glu, resulted in a virus, AV18, which was unable to transmit. In a recent work[21], the authors demonstrated that SC18 HA showed highly specific and high affinity binding to human receptors, while AV18 HA

demonstrated high affinity binding to avian receptors. NY18 bound to both avian and human receptors at a substantially lower binding affinity than that observed for AV18 and SC18, respectively. Despite dramatic differences in the aerosol transmissibility of these three viruses and their glycan receptor binding properties, they notably showed similar infectivity, replication efficiency and tissue distribution in ferrets directly inoculated with virus. More recently this approach was extended to the 2009 H1N1 system, revealing that the receptor specificity affects neither replication nor virulence of this pandemic virus in both mice and ferrets, again after intranasal inoculation, but did affect animal-to-animal transmission by respiratory droplet [25]. These studies highlight the significance of HA mutations and glycan receptor-binding specificity (given that all other genes between the three viruses are intact) in distinguishing aerosol transmissibility from other phenotypic properties of the virus such as infectivity and replication efficiency.

The dramatic changes in relative glycan receptor binding affinities and aerosol transmissibility resulting from single amino acid changes to SC18, prompted an investigation into the structural nuances governing HA-glycan interactions. While HA-glycan X-ray co-crystal structures revealed differences in overall glycan topology ('cone-like' vs. 'umbrella-like'), when bound to different HAs, these 'static' structures did not entirely capture the restrictions imposed on the conformational space of the glycan receptor by the RBS of different HAs, in moving from the free to HA-bound state.

For what is, to the best of our knowledge the first time, we present solution structures of LSTc (human receptor -Neu5Ac-α(2→6)-Gal-β(1→4)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc) and LSTa (avian receptor - Neu5Ac-α(2→3)-Gal-β(1→3)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc) bound to SC18, NY18 and AV18 HAs,

based on comprehensive NMR (STD and trNOESY) analyses and molecular dynamics simulations. The combination of both NMR and molecular dynamic simulations is particularly powerful when investigating protein: carbohydrate interactions, specifically when dealing with the highly mobile carbohydrate ligand [26]. We find that these 'dynamic' structures are instrumental in delineating the conformational features of glycans when bound to HA. Our study captures the active changes in conformation, θ and glycosidic linkage torsional angles, of LSTa and LSTc induced upon binding to HA and shed light on distinct structural constraints imposed by the RBS that differ by one or two amino acids. We report the new finding of defining the restriction on the glycan conformational space and mobility of glycan bound to HA that, importantly, is not captured in the corresponding 'snapshot' co-crystal structures. Importantly, these data are consistent with the observed differences in the biochemical binding affinities to these glycans to SC18, NY18 and AV18 HAs. We discuss the implications of these findings for our understanding of HA binding to human receptors specifically the significance of glycan conformation, θ angle and glycosidic torsional angles, and HA RBS interactions in determining appropriate specificity and affinity. Our findings will greatly aid future studies aimed at delineating appropriate structural constraints on glycan topologies for other influenza subtypes (such as H5N1 and H7N9), including surveillance of adaption to humans by these deadly viruses.

RESULTS AND DISCUSSION

NMR Structural analyses of HA-glycan Interactions. NMR analyses were performed using a 900 MHz spectrometer, which permitted unprecedented resolution when observing HA-glycan interactions. Saturation-transfer

difference (STD) NMR experiments were employed to obtain a qualitative description of the glycan residues that interact with HA. Nuclear magnetic resonance analysis indicated that for both LSTa and LSTc (Fig. 1) in complex with the different HAs, the main sugar residue involved in the contact is the terminal sialic acid (Neu5Ac). However, there are differences in the mode of binding for these glycans to the various HAs.

To aid assignment of the 1D STD signals spectra of LSTc bound to SC18 and NY18 were superimposed on the HSQC spectrum of LSTc (Fig. 2 and 3). While the interaction between LSTc and SC18 occurs primarily through the non-reducing end Neu5Ac moiety, involving protons: H3ax, H3eq, H4, H5, H7 and H9 (Fig. 2), signals belonging to Gal-1 (H6), GlcNAc (methyl group), Gal-2 (H4 and H1) and Glc (H5, H6) are also present in the STD spectrum (Fig. 2).

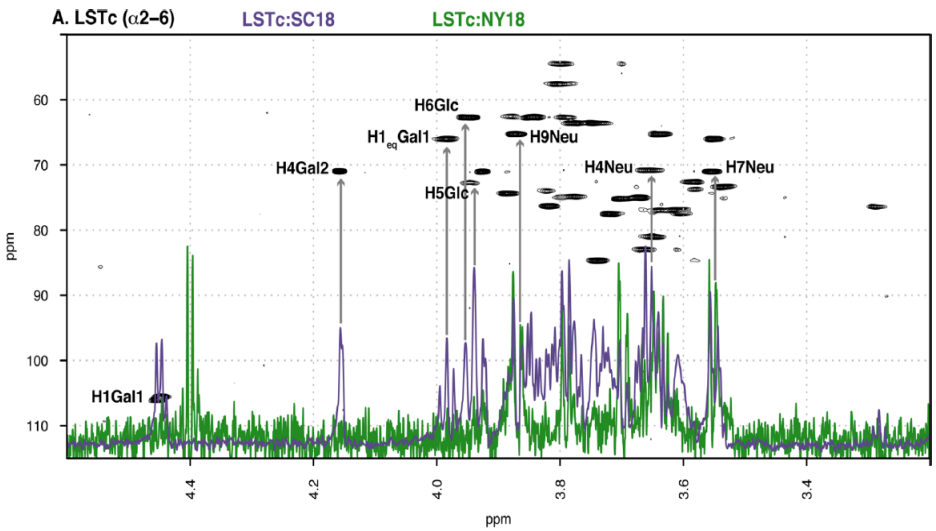


Fig. 2 Main region of the overlaid STD HSQC spectra of LSTc receptor complex. STD spectra of LSTc-SC18 (purple) and LSTc-NY18 (green) complexes overlaid upon the HSQC spectrum of LSTc.

Similarly, LSTc interacts with NY18 principally through Neu5Ac, (H4, H5, H7 and H9), whereas, signals belonging to Gal-2 and Glc appear to be absent, or at best, weaker compared to those found in the STD spectrum of LSTc-SC18.

Notably, both methyl groups of GlcNAc and Neu5Ac still interact with NY18, but these signals are weaker with respect to those observed in the LSTc-SC18 complex (Fig. 3).

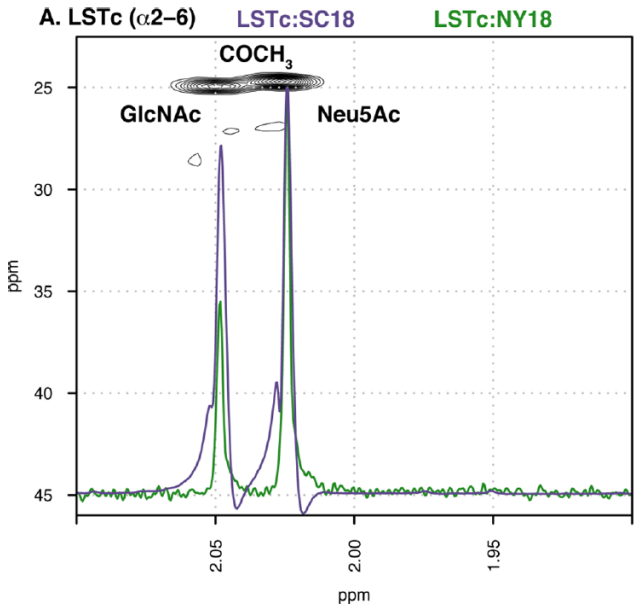


Fig. 3. N-Acetyl regions of the overlaid STD HSQC spectra of LSTc-receptor complex. STD spectra of LSTc-SC18 (purple) and LSTc-NY18 (green) complexes overlaid upon the HSQC spectrum of LSTc.

The greater number of proton signals corresponding to the monosaccharides of LSTc that interact with SC18 compared to NY18 is consistent with the dose-dependent direct biochemical binding of these HAs to this glycan in an array platform (Fig. 4) and also with previously reported higher human receptor

affinity of SC18 HA over NY18 HA [21]. The absence of STD signals observed for the complex LSTc-AV18 (Fig. 5), indicates that LSTc, does not interact with AV18, consistent with minimal to no binding of AV18 to LSTc that is observed in the dose-dependent binding assay (Fig. 4).

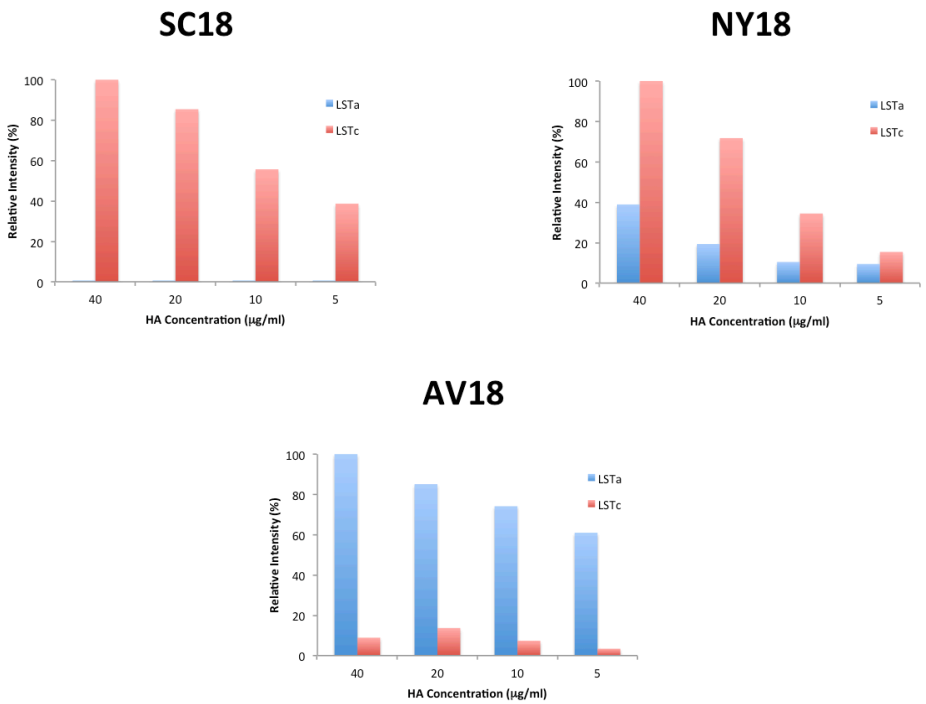


Fig. 4 Dose-dependent binding of SC18, NY18 and AV18 Haemagglutinin to LSTa and LSTc.

The avian receptor, LSTa, on the other hand, interacts with AV18 and NY18 almost exclusively through Neu5Ac (Fig. 6). Owing to the overlapping signals of the methyl groups belonging to Neu5Ac and GlcNAc it was not possible to establish definitively which group interacts with the corresponding HA (Fig. 7). Among the few STD signals that were observed for the interaction between LSTa-AV18/NY18 that did not belong to Neu5Ac, weak signals were observed

for Gal-1 H3 and H4, consistent with the partial involvement of Gal-1 residue in the binding.

Given the conformational flexibility of glycans, arising from the numerous glycosidic torsion angles, the parameter θ has been defined to quantify the form of the non-reducing end of the receptors (Fig. 1 and Methods Section). To compare the conformational space sampled by the free and bound ligand, tr-NOESY experiments were performed for LSTc interacting with SC18 and NY18 (Fig. 8).

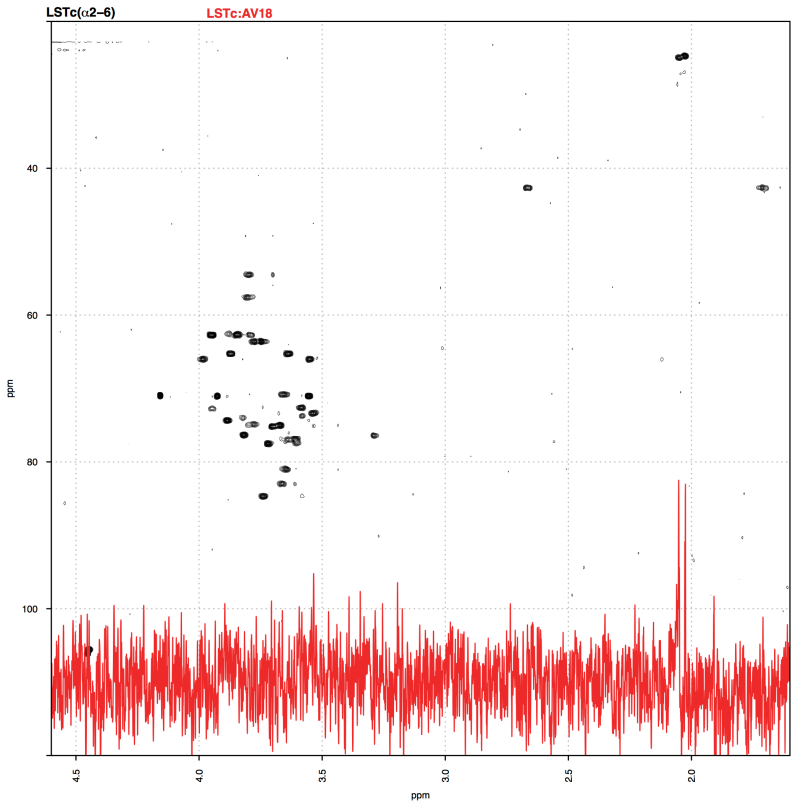


Fig. 5 LSTc:AV18 STD spectrum (red) superimposed on the top of the HSQC spectrum of LSTc (black). There is little evidence of an interaction between the receptor and HA.

Notably, by comparing the results from bound and unbound LSTc, the NOE signals of the GlcNAc methyl group are substantially different.

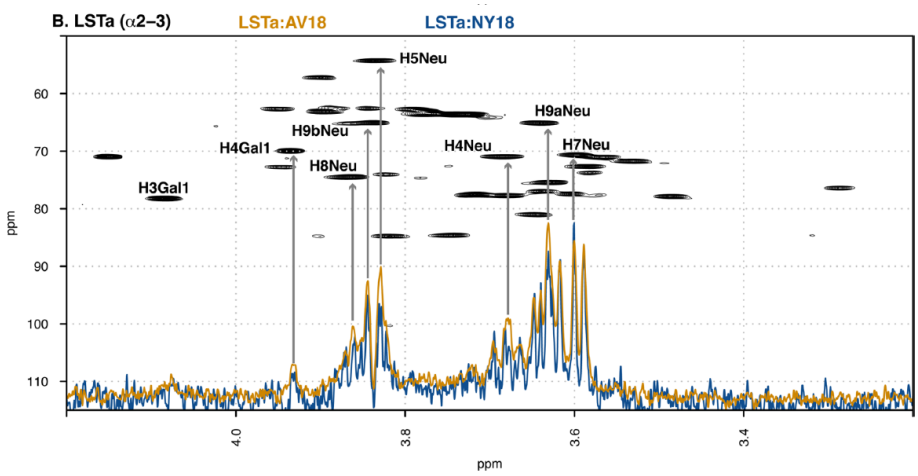


Fig. 6 Main region of the overlaid STD HSQC spectra of LSTa-receptor complexes. STD spectra of LSTa-AV18 (orange) and LSTa-NY18 (blu) complexes overlaid upon the HSQC spectrum of LSTa.

Whereas, only the NOE between the methyl group of GlcNAc and H5 of Neu5Ac was observed for unbound LSTc, additional NOEs are observed in the bound state with SC18, including signals associated with the H8/H9 protons. These findings agree with a reduction in the distance between the GlcNAc (CH3 protons) and Neu5Ac (H5, H8/H9 protons), which is likely correlated with a reduction of the θ angle between the free and the bound state.

This analysis indicates that, upon binding to SC18 and, to a lesser extent NY18, LSTc undergoes a conformational change, reducing the θ angle and subsequent conformational freedom of the Neu5Ac residue, particularly in the region between carbons C6 to C9.

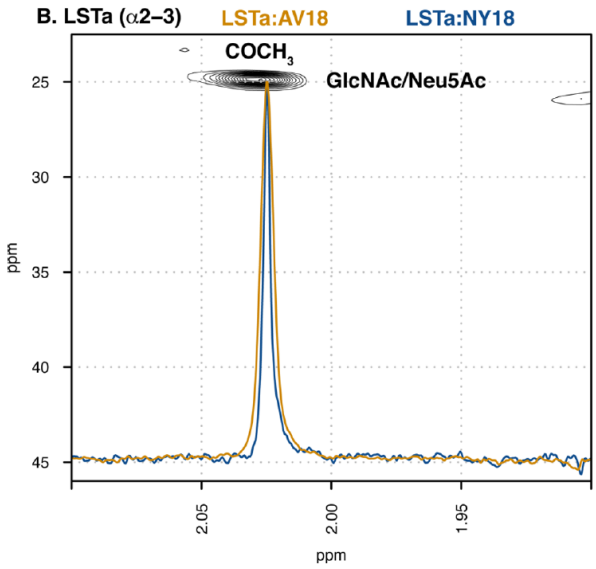


Fig. 7 N-Acetyl region of the overlaid STD HSQC spectra of LSTa receptor compexes. STD spectra of LSTa-AV18 (orange) and LSTa-NY18 (blu) complexes overlaid upon the HSQC spectrum of LSTa.

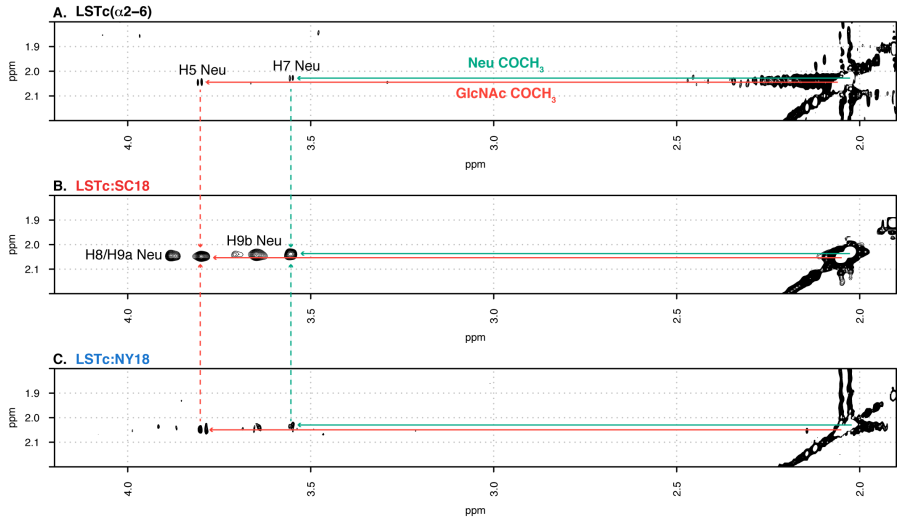


Fig. 8 NOESY spectrum of LSTc (A.) and tr-NOESY spectra of LSTc-SC18 (B.) and LSTc-NY18 (C.). The new stronger dipolar interactions measured between the CH3 of GlcNAc and Neu5Ac (H5, and H8/H9) residues suggests a reduction of the distances between these contacts, corresponding to a decrease in the θ angle induced by binding to HA, and a reduction of the Neu5Ac sialyl (C6 to C9) conformational freedom. The change in NOE magnitude is more significant in the case of SC18 than it is for NY18.

Molecular Dynamic simulation of HA-glycan interactions. The impact of glycan conformational differences and points of glycan-HA interactions observed in the NMR experiments were further investigated using MD simulations, allowing a structural and dynamic comparison between the various complexes (LSTc-SC18/NY18/AV18 and LSTa-AV18/NY18) to be made.

During the MD simulations the conformational and dynamic properties of the HA-glycan complexes progressively change, especially the glycan conformation, with differences arising in the complexes due to the amino acid mutations within the receptor-binding site (RBS). It should be noted that the glycan starting geometries are the same in the two sets of MD simulations, with the LSTa and LSTc starting geometries taken from the solution forms identified during the previous work with the glycan free in solution[27]. The mobility observed for the glycan in the HA RBS was monitored using principal component analysis (PCA). Briefly, the analysis was performed on the distance matrix between the HA and receptor, unlike the conventional PCA of protein MD simulations which uses the position of the protein backbone or the dihedral angles of the protein, this focuses the analysis on the interaction between receptor and ligand.

The analyses were performed on the time interval 20 – 120 ns, the MD simulation trajectories were decomposed in to distance matrices between the glycan atoms and HA amino acid side chains, it was these distance matrices that were scrutinized by PCA.

The temporal changes in the conformational state of LSTc-SC18 are reported in Fig. 9C, the conformer subset identified by cluster g3 represents the final bound state of the complex, a similar representation for LSTc-NY18 and LSTc:AV18 is shown in Fig. 9F and J.

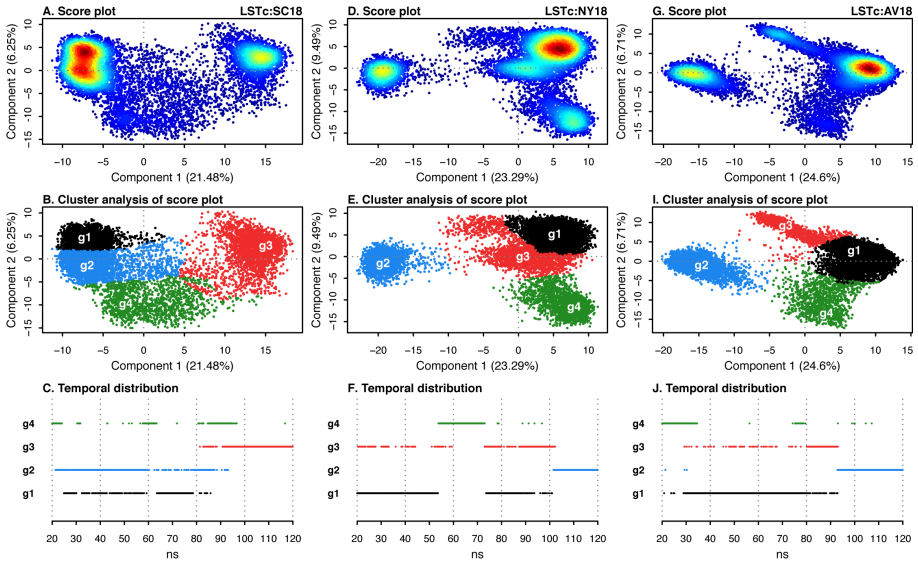


Fig. 9 *PCA analysis of the LSTc bound to SC18 (**A**. - **C**.), NY18 (**D**. - **F**.) and AV18 (**G**. - **J**.)*. Panels **A**., **D**. and **G**. contain the score plots produced by PCA, the clusters within this plot represent the different glycan-HA conformational states visited in the MD simulation. The conformational subsets were identified by performing non-parametric density estimation cluster analysis on the score plot (**B**., **E**. and **I**.). The occurrence of the different conformational subsets is time dependent (**C**., **F**. and **J**.), with the MD simulations ending with each of the LSTc-HA complexes finding a single, final, state. The colour represents the density of conformers in each state, red represents a high density while blue signifies low density.

When LSTc is bound to any of the HAs it appear to find a final singular state, this is not the case for LSTa-AV18 or LSTa-NY18, the final state of the former MD simulation is represented by at least 4 conformational subsets (Fig. 10C; clusters g2, g3, g4 and g5), while the latter has 2 final conformational subsets (Fig. 10F; clusters g1 and g2), indicating a greater conformational freedom compared to LSTc interacting with HAs.

Inspection of the MD trajectory indicates that, in the case of LSTc-SC18, all the monosaccharides of LSTc are positioned to interact with the RBS, this could be visualized in Fig. 11A, which describe the contact network between LSTc and SC18.

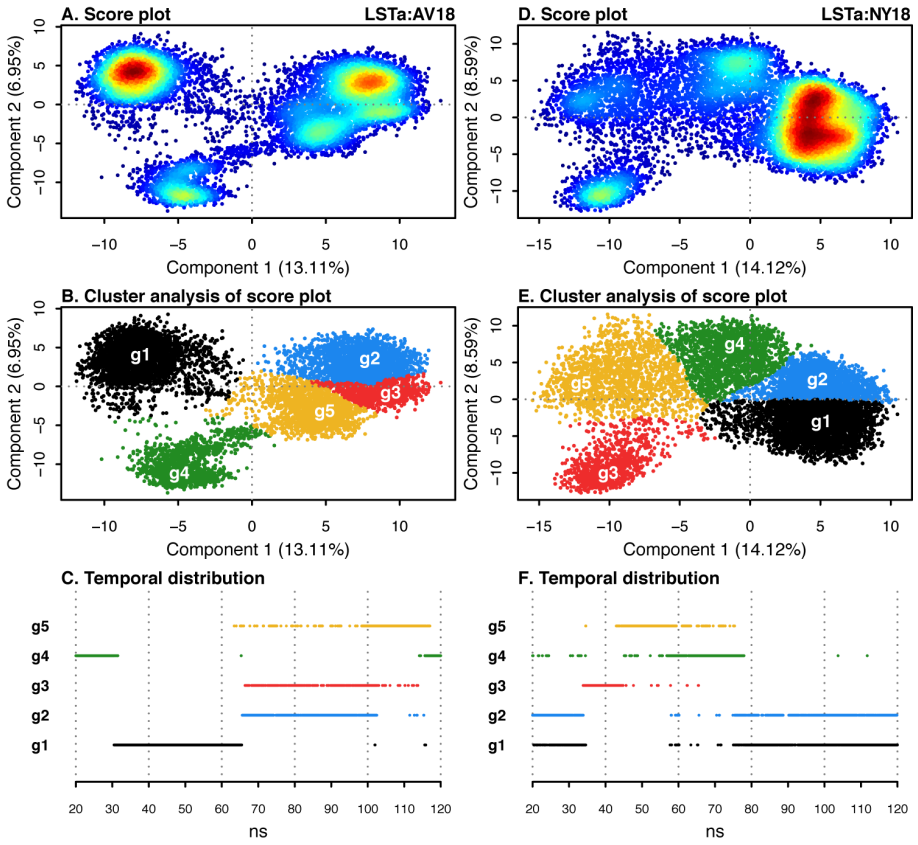


Fig. 10 . *PCA analysis of the LSTa bound to AV18 (**A**. - **C**.) and NY18 (**D**. - **F**.)*. Panels **A**. and **D**. contain the score plots produced by PCA, the clusters within this plot represent the different glycan:HA conformational states visited in the MD simulation. The conformational subsets were identified by performing non-parametric density estimation cluster analysis on the score plot (**B**. and **E**.). The occurrence of the different conformational subsets is time dependent (**C**. and **F**.), with the MD simulations ending with each of the LSTa-HA complexes have many states at the end of the MD simulation, indicating that the receptor is very mobile in the RBS. The colour represents the density of conformers in each state, red represents a high density while blue signifies low density.

It is also apparent that both hydrogen bonds and dispersive forces are important components in the interaction of LSTc with SC18, specifically the methyl groups of GlcNAc show persistent contacts with Leu194 (Fig. 11D) and Asp190, while the methyl group of Neu5Ac interaction with Gly134 and Trp153

(Fig. 11A). Another noteworthy interaction within the LSTc-SC18 complex is between Gal-1 and Gln226, Lys222 and Asp225. The single-point mutation forming NY18, Asp225-Gly, drastically alters the interaction between the non-reducing end of the receptor and HA, and the major interactions with Gal-1 are removed (panel 11B compared to panel 11A). The presence of Gly instead of Asp in position 225, as a matter of fact, removes the hydrogen bond interaction between Gal-1 (OH3, OH4) and the RBS (specifically involving residues Gln226, Lys222 and Asp225) (Fig. 11B), thereby promoting a greater distance between Gal-1 and HA (Loop220) compared to SC18 (Fig. 12B and Fig. 11B).

This result is clearly supported by the STD NMR data (Fig. 2 and 3), where Gal-1, GlcNAc, Gal-2 and Glc resonances are weaker or absent in the spectrum of LSTc-NY18 compared to the wild type complex, corresponding to a weaker interaction. The additional modification forming AV18, Asp190-Glu, allows Gal-1 to re-engage with HA, which leads to the interaction between GlcNAc and amino acids 190 and Ser193 being lost (compare panel 11C to panels 11B and 11A). In fact, the LSTc-AV18 complex the further mutation of Asp19-Glu, introduces a greater steric hindrance preventing the optimal interaction between the reducing end of LSTc and Helix190 of AV18, as can be seen by comparing the corresponding distances in the MD simulation trajectories of LSTc-AV18 with the reference complexes LSTc-SC18 or LSTc-NY18 (Fig. 11).

The consequence of this is that the interaction between Gal-1 and the HA RBS is re-established and Neu5Ac is drawn closer to the RBS, while the interaction of GlcNAc is weaker compared to LSTc-SC18 and LSTc-NY18 (specifically Asp190 and Leu194 with GlcNAc-OMe) (Fig. 11D), with the whole residue moving away from the RBS.

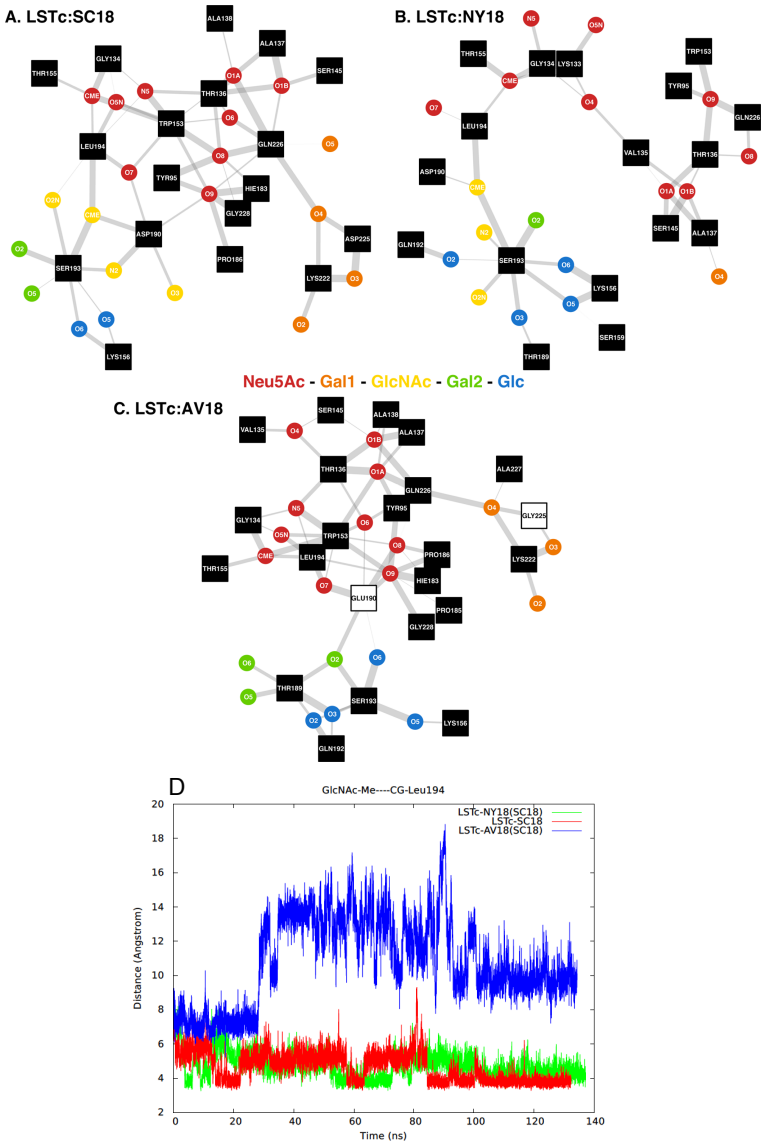


Fig. 11 (A-C) Contact network between LSTc and HA. If two vertices are linked, they are <6 Å apart; the thicknesses of the graph edges are inversely proportional to the distance between the glycan atom (circular vertex) and the protein amino acid (square vertex) (the thicker the edge, the closer the two are together). The square vertices that are white with a black boundary indicate that this amino acid has been mutated. The networks represent the average distances found in the final, "bound", conformer subsets identified by PCA (Fig. 9): cluster g3, LSTc−SC18; cluster g2, LSTc−AV18; and cluster g2, LSTc−NY18. (D) Distances between Leu194 and CH_3 of GlcNAc of SC18(red), NY18(green) and AV18(blu) during the MD simulation.

The analyses here reaffirms that it is interaction of Neu5Ac, Gal1 and GlcNAc with HA that is important for human adaptation of influenza.

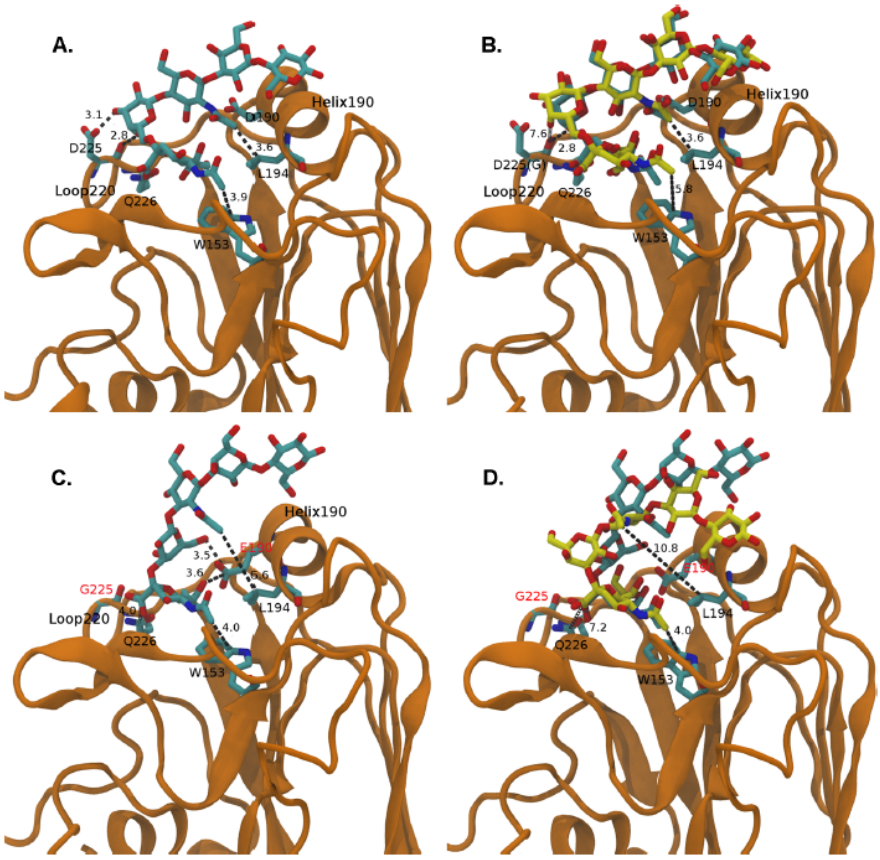


Fig. 12 (A) Structure of the LSTc−SC18 complex in the g3 cluster conformation subset (Figure 9C). (B) Superposition of the RBSs of the LSTc−SC18 complex with that of the LSTc−NY18 complex (rmsd = 0.78 Å). The reported complex structures belong to g3 and g2 conformations, respectively (Fig. 9C and F). The carbon skeleton of LSTc bound to SC18 is colored cyan, while LSTc interacting with NY18 is colored yellow. (C) Structure of the LSTa−AV18 complex corresponding to conformation subset g2 of the PCA conformational characterization (Fig. 10C). (D) Superimposed structures of the LSTa−AV18 and LSTa−NY18 complexes (rmsd = 0.74 Å). The reported structure of the LSTa−NY18 complex corresponds to g2 and g1 conformation subsets as obtained from PCA (Fig. 10F). LSTa linked to AV18 is colored cyan, while LSTa interacting with NY18 is colored yellow. The relevant amino acid residues of the shown HA active sites are underlined by a tube representation, with the name and numbering relative to PDB entry 2WRG. The reported distances are in angstroms.

These results provide a structural description of the effect caused by the single and double amino acid mutation in the RBS of HA, which correlates with a progressive weakening of the interaction between LSTc and HA (SC18 > NY18 > AV18), a result also supported by NMR-STD experiments and the measured binding affinities.

In the case of the LSTa-AV18 complex, the primary contacts involve Neu5Ac of LSTa, with the carboxyl-, acetyl- of Neu5Ac interacting with the HA RBS (Fig. 12C and Fig. 13). Importantly, contacts are also observed between AV18 and Gal-1 in LSTa. This involves the interaction of O6-Gal-2 with Glu190 and Pro186, which is afforded by the presence of the α(2→3) linkage in LSTa. The PCA of the LSTa-AV18 MD simulation trajectory also indicates a greater mobility of the reducing end region of LSTa compared to its non-reducing terminal (Neu5Ac), which remains strongly attached to the RBS during the MD simulation time scale (Fig. 10, and Fig. 13). Comparing LSTa-NY18 with LSTa-AV18 complexes, the mutation of Glu190-Asp reduces the interaction between Gal-1 and the RBS, with only the non-reducing end Neu5Ac residue interacting with the RBS. A consequence of this is that the number of interactions between the Neu5Ac residue and the RBS are decreased and the interaction between Gln226 and the carboxyl group of Neu5Ac is no longer observed (Fig. 12D, and Fig. 14).

This corresponds to a weaker interaction between LSTa and NY18 compared to AV18, in agreement with the NMR-STD experiments and biochemical assay. The LSTa-NY18 complex samples two conformational states during the MD simulation, which differ based on contacts between Neu5Ac and the HA RBS (Fig. 14).

The features of LSTa binding to AV18 and NY18 obtained from the MD simulations is consistent with the corresponding NMR STD signals, where data

from both complexes indicate that Neu5Ac is the main interacting residue with HA, while the methyl group of GlcNAc is no longer a key point of interaction with either NY18 or AV18.

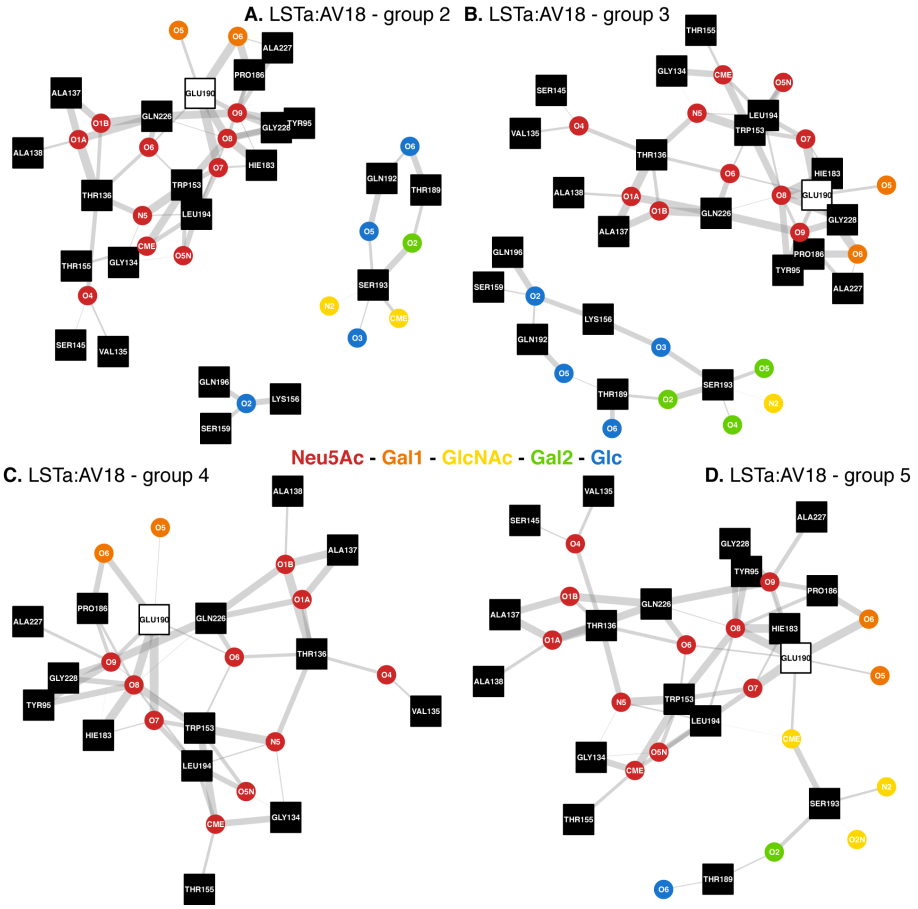


Fig. 13 Contact network between LSTa and AV18. If two vertices are linked they are less than 6 Å apart, the thicknesses of the graph edges are proportional to the distance between the glycan atom (circular vertex) and protein amino acid (square vertex), the thicker the edge the closer the two are together. The square vertex that are white with a black boundary indicate that this amino acid has been mutated, for example Asp190→Glu. The networks represent the average distances found in the final, 'bound', conformer subsets identified by PCA, Fig. 10 panel c., clusters g2, g3, g4 and g5.

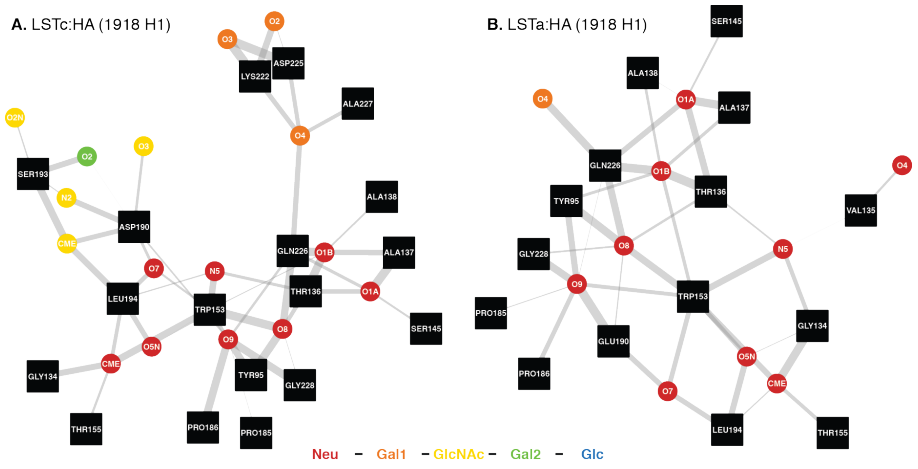


Fig. 14 Contact network between LSTa and NY18. The networks represent the average distances found in the final, 'bound', conformer subsets identified by PCA, Fig. 10 panel f., clusters g1 (left) and g2 (right).

Dynamics of LSTa and LSTc conformation upon binding to HA. The θ angle parameter is a key conformational descriptor of the non-reducing end of the glycan receptor, indicating the different forms of the glycans. Here we analyzed the θ angle parameter for the PCA-extracted conformers. As shown previously, the θ angle of unbound LSTc had a predominant distribution at 87° and a smaller population located at 119° (Fig. 15A) [27]. Binding to SC18 substantially restricts the conformational population of LSTc, as reflected by the narrow distribution of the θ parameter, with a single distribution centered at 82°. Binding to NY18 also restricts the conformational space sampled by LSTc, with a θ angle distribution being located at approximately 90°. Interestingly, the small population of θ angles values around 119° in the unbound LSTc is absent in the bound state, consistent with an earlier study in which it was postulated that

long α(2→6) glycans would predominantly adopt umbrella-like topology (characterized by θ < 100°) when bound to the RBS of human-adapted HAs [19].

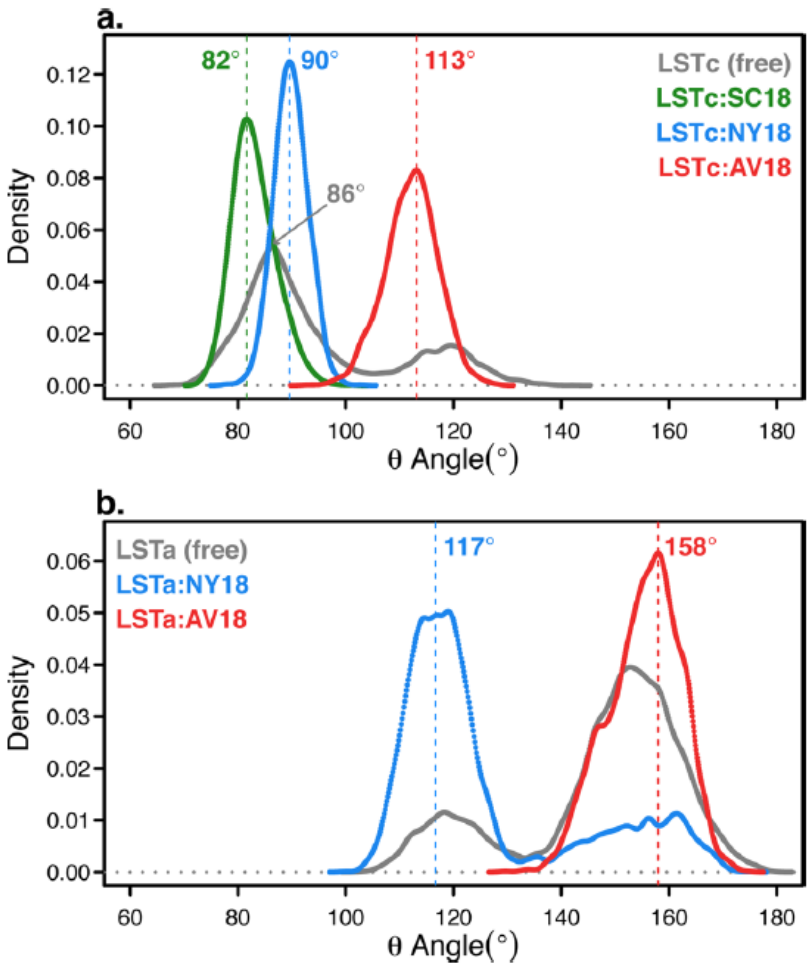


Fig. 15 Density distribution plots of the topological θ angles of LSTc (A) and LSTa (B) free and bound to HA. The interaction of LSTc or LSTa with HA alters the topological θ angle assumed by the glycan receptor. The θ angles shown here are for the PCA-extracted conformers; a comparison of these with the θ angles for all of the conformers can be found in Figure 16.

This restriction in θ angle when LSTc binds to SC18 and NY18 is experimentally supported by the new NOE signal appearing between GlcNAc and Neu5Ac

H8/H9 in the tr-NOESY spectra of LSTc-SC18 and by the stronger NOE signal between GlcNAc and Neu5Ac H5 in LSTc-NY18 (Fig. 8). In contrast, low affinity interaction of LSTc with AV18 predominantly samples cone-like topology as indicated by the θ distribution around 113°, which is consistent with the constraints imposed by avian-adapted HAs and is due to the lack of interaction between GlcNAc and the RBS. These results show how LSTc assume distinct conformations originated by the different constrains imposed by the three HA: SC18, NY18 and AV18, corresponding to different biochemically measured affinity.

In the unbound state the θ angle of LSTa samples two distinct populations centred around 118° and 154° (Fig. 15B), both of which correspond to 'cone-like' topologies (θ > 100°). The binding of LSTa to NY18 does not significantly alter the position of the θ angle distribution compared to free glycan, distributions located at 117° and 158° in the bound state, but it does strongly affect their relative populations, with the population located at the smaller θ angle being the largest. The binding of LSTa to AV18 HA imposes restrictions on the θ angle, removing the population distributed around 117° (Fig. 15B). Consistent with our previous studies, the avian receptor LSTa exclusively adopts a 'cone-like' topology whether present in the unbound or when bound to NY18 and AV18 HA. For completeness the θ angle density plots for the entire MD trajectory [20 – 100 ns] can be found in Fig. 16.

To complete our conformational studies we also considered the glycosidic torsion angles of the free and bound glycan receptor sampled during the MD simulation - in the case of the bound receptors the conformer subset extracted by PCA are considered (Tables 1 and 2).

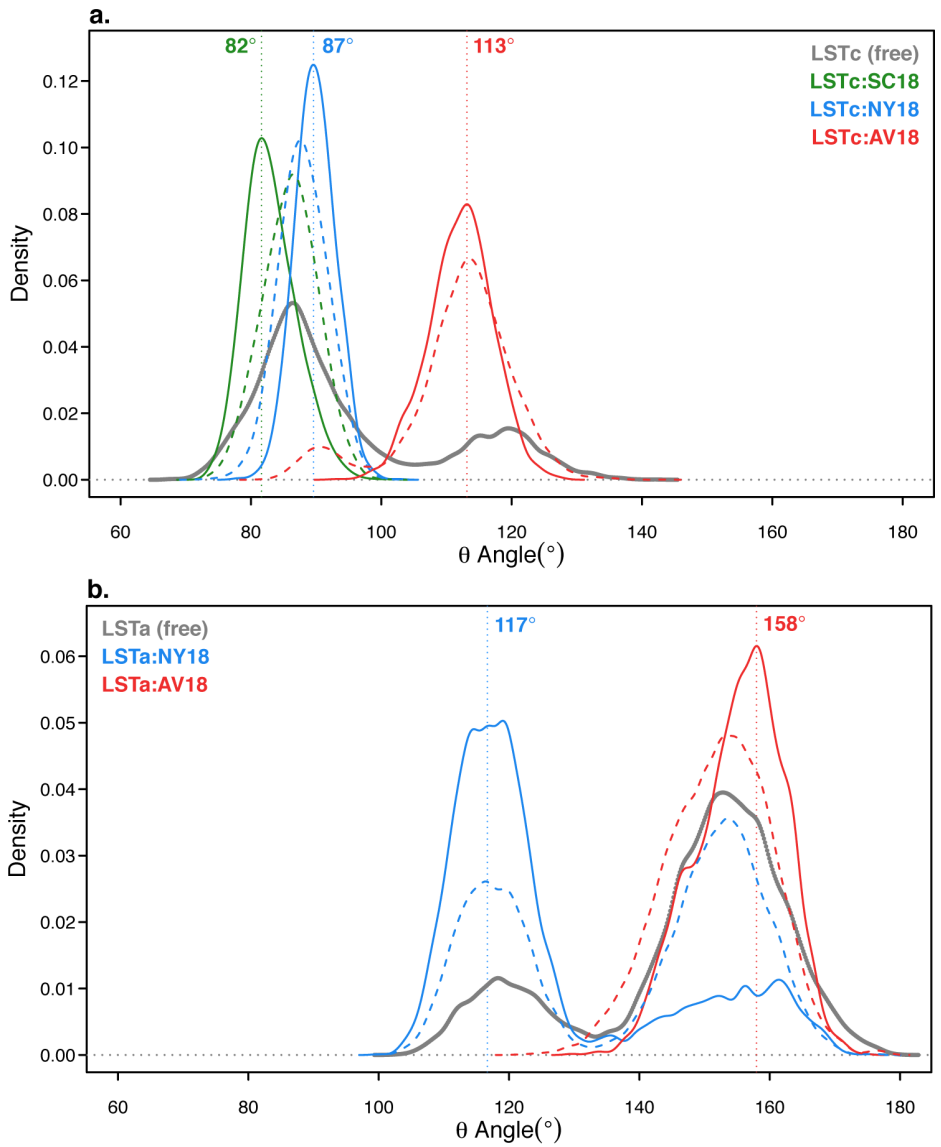


Fig. 16 θ angle observed for LSTx when free and bound to HA (A. LSTc-HA and B. LSTa-HA). The θ angles seen in the period 20-120 ns of the MD simulations (dashed lines) are compared to those seen in the PCA selected conformers (solid lines) [LSTc PCA selected conformers: cluster g3 LSTc-SC18, cluster g2 LSTc-AV18 and cluster g2 LSTc-NY18 (Fig. 9) and LSTa PCA selected conformers: LSTa-NY18 clusters g1 and g2 and LSTa-AV18 clusters g2, g3, g4 and g5 (Fig. 10)].

First, consistent with the θ angle distributions described above, comparison of the glycosidic torsional angles in the unbound and various HA-bound glycans indicates there are distinct structural constraints imposed by the RBS of SC18, NY18 and AV18 HA on LSTa and LSTc.

	Linkage	Cluster	Size	Percentage	Phi ° (ϕ)	SD Phi ° (ϕ)	Psi ° (ψ)	SD Psi ° (ψ)
LSTc	1	1	9858	98.6	-57	11	190	17
		2	48	0.5	-71	12	-114	5
		3	73	0.7	-60	9	111	13
	2	1	5430	54.3	42	11	-3	11
		2	2419	24.2	-38	13	-29	11
		3	2151	21.5	24	13	-32	10
	3	1	4691	46.9	20	14	-47	13
		2	3981	39.8	39	15	34	18
		3	1328	13.3	-25	11	-43	11
	4	1	5934	59.3	39	13	-8	18
		2	4066	40.7	-37	14	-30	13
LSTc:SC18	1	1	2763	100.0	-58	7	189	7
	2	1	2763	100.0	50	9	-5	9
	3	1	2724	98.6	-2	21	-42	11
		2	39	1.4	40	7	25	8
	4	1	1192	43.1	45	10	0	11
		2	1069	38.7	19	13	-30	11
		3	502	18.2	-33	18	-28	11
LSTc:NY18	1	1	1826	100.0	-46	10	194	7
	2	1	1826	100.0	41	9	-7	11
	3	1	1826	100.0	17	14	-43	12
	4	1	1826	100.0	-31	13	-27	9
LSTc:AV18	1	1	2698	100.0	-51	8	192	7
	2	1	2698	100.0	-24	14	-32	11
	3	1	2698	100.0	24	13	34	11
	4	1	2698	100.0	-28	12	-24	9

Table 1. Glycosidic torsional angles for LSTc, free and bound to SC18, NY18 and AV18. The angles presented in the table for the ligand receptor complexes were determined for those conformers that were extracted by PCA; g3, LSTc-SC18; g2, LSTc-NY18 and g2, LSTc-AV18 (Fig. 9). The angles were determined by cluster analysis of the data illustrated in Fig.16, the analysis used nonparametric density estimation to determine the members of each cluster. The angles represented in the table are the average values for each ϕ and ψ cluster.

When LSTc binds to SC18 and NY18 the distribution of states seen for φ1/ψ1, φ2/ψ2 and φ3/ψ3 are very similar, only by using density cluster analysis is it

possible to observe the different central locations of the states (Fig. 17 and Table 1).

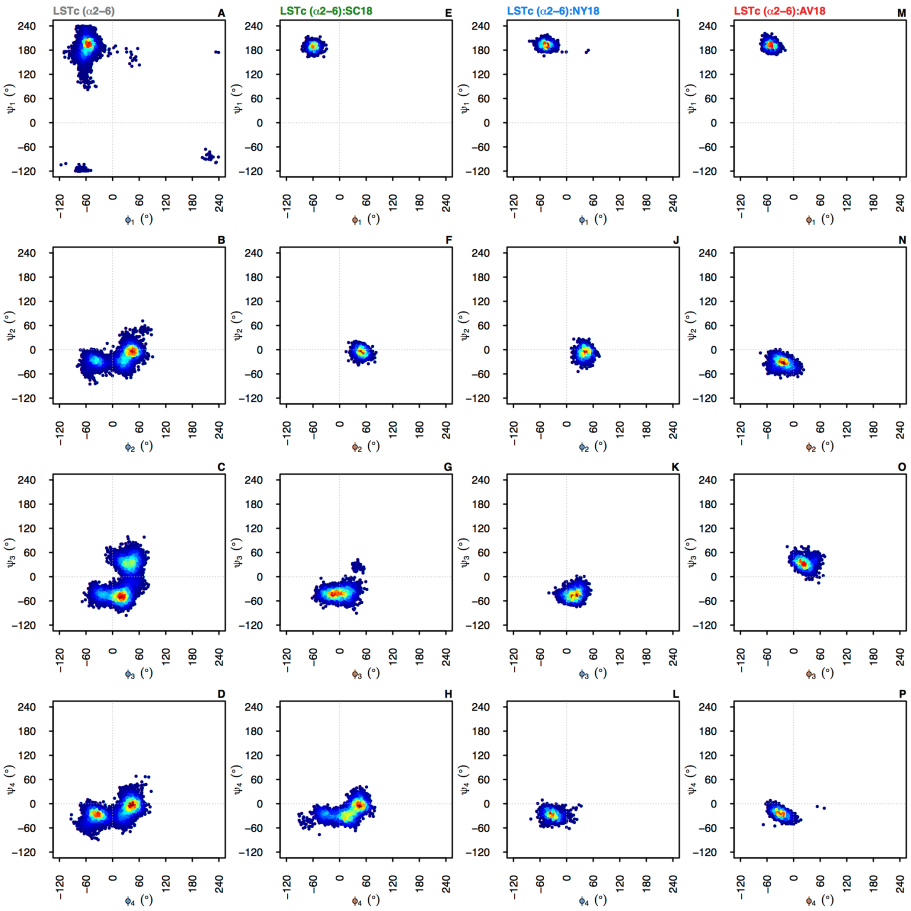


Fig. 17 Glycosidic torsional angles of LSTc, free (A - D) and when bound (PCA selected conformers; g3, LSTc-SC18; g2, LSTc-NY18; g2, LSTc-AV18 (Fig. 9 and Fig. 11)) to SC18 (E - H), NY18 (I - L) and AV18 (M - P). The colour represents the density of conformers in each state, red represents a high density while blue signifies low density.

Indicating the modification of amino acid 225 (Asp-Gly) and the subsequent change in the interaction between LSTc and HA, the loss of the interaction of Gal-1 with Gln226, Asp225, and Lys222 does not have a strong effect on the

glycosidic torsional angles, whereas binding of LSTc to AV18 differing from SC18 in position 190 and 225, which eliminates the interaction between GlcNAc and Leu194, Asp190, and Ser193, affecting the positions of φ2 and ψ2 and those of φ3 and ψ3, which are distinct from those observed in the LSTc−SC18 and LSTc−NY18 complexes (Figure 17 and Table 1).

In the case of when LSTa binds to HA, the interactions with AV18 and NY18 are very different. The interaction between LSTa and NY18 is solely through the non-reducing end Neu5Ac (Fig. 14) and this is evident by the diversity of states observed for φ2/ψ2, φ3/ψ3 and φ4/ψ4 (Fig. 18 and Table 2).

The opposite is seen for the LSTa-AV18 complex, even though the glycan is dynamic with four conformational states being extract by PCA of the complexes MD trajectory (Fig. 10), the glycosidic torsional angle states observed in the glycan are restricted for all four linkages.

CONCLUSION

The NMR and MD simulation analyses described in this study offer new insights into the interaction between haemagglutinin and its glycan receptors, providing a detailed description of the contacts observed in the interactions between LSTc and LSTa withH1 of Sc18 and their tested mutants, and their consequential change in glycan conformation. The principal consequence of modifying SC18 to form NY18, Asp225-Gly, is that Gal-1 of LSTc can no longer interact with the HA RBS (Gln226, Lys222 and Asp225), which allows Gal-1 to move away from the protein surface, also effecting the interaction between the Neu5Ac residue of LSTc and the RBS.

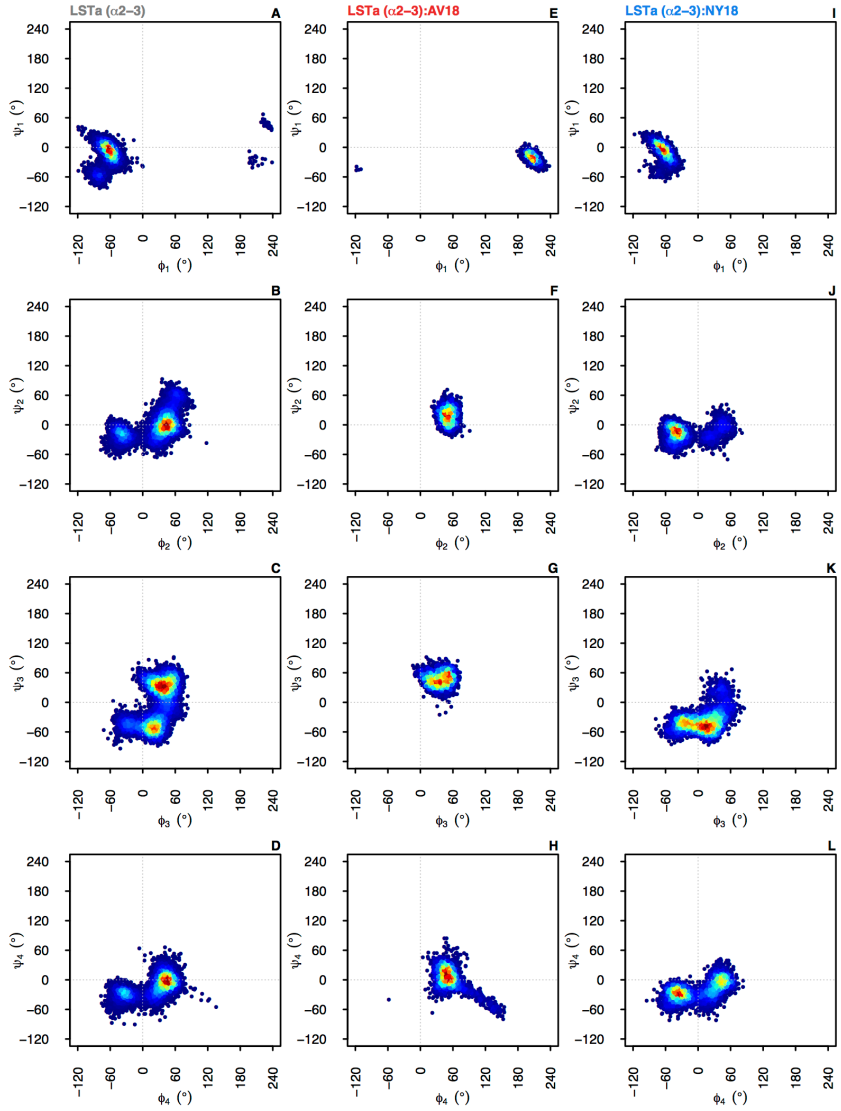


Fig. 18 Glycosidic torsional angles of LSTa, free (A-D) and when bound (PCA selected conformers) to AV18 (E-H) and NY18 (I-L). The colour represents the density of conformers in each state; red represents a high density while blue signifies low density. The torsional angles of LSTa bound to AV18 and NY18 are the combination of the states found by PCA to be the final 'bound' state (AV18 (g2/g3/g4/g5) – Fig. 10 and 13- NY18 (g1/g2) Fig. 10 and 14). A trend can be seen that as the affinity between the receptor, LSTa, and the various ligands decreases the variation in the glycosidic torsional angles states increases. In the case of LSTa-NY18 the variation in the reducing end linkages is similar to that observed in the free receptor.

	Linkage	Cluster	Size	Percentage	Phi ° (ϕ)	SD Phi ° (ϕ)	Psi ° (ψ)	SD Psi ° (ψ)
LSTa	1	1	8771	87.7	-62	10	-7	13
		2	1191	11.9	-81	9	-55	9
		3	21	0.2	230	6	47	8
		4	17	0.2	209	12	-26	8
	2	1	7317	73.2	41	13	-1	17
		2	1917	19.2	-37	13	-20	13
		3	766	7.7	61	12	57	10
	3	1	5939	59.4	39	16	32	19
		2	3290	32.9	20	13	-48	14
		3	771	7.7	-28	11	-44	13
	4	1	7861	78.6	40	13	-6	17
		2	2139	21.4	-34	14	-30	12
LSTa:AV18	1	1	2426	100	206	15	-20	8
	2	1	2426	100	51	9	19	15
	3	1	1118	46.1	49	9	53	13
		2	1308	53.9	25	12	42	12
	4	1	2426	100	54	22	5	21
LSTa:NY18	1	1	3421	97.1	-64	11	-6	14
		2	101	2.9	-73	10	-50	8
	2	1	2674	75.9	-40	11	-16	14
		2	484	13.7	45	11	1	12
		3	364	10.3	23	13	-26	10
	3	1	1939	55.1	22	15	-44	14
		2	1234	35	-24	17	-42	14
		3	349	9.9	47	14	14	19
	4	1	1760	50	-36	14	-30	12
		2	1762	50	35	15	-11	19

Table 2. Glycosidic torsional angles for LSTa, free and bound to NY18 and AV18. The angles presented in the table for the ligand receptor complexes were determined for those conformers that were extracted by PCA; g2/g3/g4/g5, LSTa-NY18; g1/g2, LSTa:AV18 (Fig. 10). Unlike when LSTc is bound to HA multiple bound states were evident when LSTa was bound to HA. The angles were determined by cluster analysis of the data illustrated in Fig. 18.The angles represented in the table are the average values for each ϕ and ψ cluster.

A further second modification forming the avian-adapted AV18, Asp190→Glu, permits Gal-1 of LSTc to re-engage with the HA RBS, interacting with Gly225 and Lys222. While, GlcNAc can no longer interact with the RBS in the LSTc-AV18 complex, it is this interaction that is a key decider for human-adaption. The interaction between LSTa and AV18, avian-adapted HA, is principally via

Neu5Ac and Gal-1 of the receptor. The non-reducing end α(2→3) linkage between Neu5Ac and Gal-1 in LSTa permits Glu190 and Pro186 to interact with the methyl-group of Neu5Ac and O6-Gal-1 of LSTa.

The interaction between the glycan receptors and HA imposes conformational constraints upon the glycan, which is characterized in terms of θ angle and the glycosidic torsional angles. We also correlated this analysis with biochemical HA-glycan binding specificity and affinity to ensure that this analysis is consistent with available crystal structure information. Importantly, using the θ angle as a parameter to characterize the overall shape of the glycan, our study demonstrates key differences in the form of LSTa and LSTc when they are free versus their bound states (LSTc-SC18/NY18/AV18 and LSTa-AV18/NY18), which have not been possible to 'capture' through x-ray crystal structures.

Previously, we have noted that human receptors in their unbound state sample a conformational space that resembles both a 'cone-like' (θ ≥ 100°) and 'umbrella-like' (θ < 100°) topology, whereas, avian receptors exclusively sample 'cone-like' topology[27]. By analysing HA-glycan co-crystal structures, we postulated that glycans binding to 'avian-adapted' HA RBS (such as AV18) would impose constraints on the human receptor (for example LSTc) to preferentially sample a 'cone-like' topology, whereas, glycans binding to 'human-adapted' HA (such as SC18) would constrain the glycan to preferentially sample 'umbrella-like' topology [19]. Based on the NMR analyses and MD simulations presented in this study, we demonstrate the former hypothesis to be correct. The θ angle of LSTc (Figure 15A) samples two distinct populations in the unbound state corresponding to 'umbrella-like' and 'cone-like' topology. Upon binding to AV18 HA, LSTc predominantly samples a 'cone-like' topology (θ ~ 110°), this is due to GlcNAc of LSTc not being able to interact with the

helix190 in the RBS of HA. On the other hand, binding to the RBS of SC18 or NY18 imposes constraints on LSTc such that the glycan exclusively samples an 'umbrella-like' topology.

Futhermore, compared to free LSTc the bound glycan has restricted glycosidic torsional angles, with the torsional states observed for LSTc-SC18 and LSTc:NY18 being similar, whereas the mutation of Asp190 to Glu, seen in AV18, which abolishes the interaction between GlcNAc of LSTc and helix190, produces unique glycosidic angles for linkage φ2/ψ2 and φ3/ψ3, which are distinct to those seen in the LSTc-SC18 and LSTc-NY18 complexes. This restriction in the glycosidic torsional angles is also observed in the LSTa-AV18 complex, even though the receptor is dynamic in the RBS, as indicated by the multiple states observed at the end of the MD simulation (Fig. 10). Whereas, in the LSTa-NY18 complex, were the non-reducing end Neu5Ac is the sole point of interaction with HA, linkages φ2/ψ2, φ3/ψ3 and φ4/ψ4 all have glycosidic torisional angles which are similar to those seen in the free glycan (Fig. 18).

This higher restriction correlates with a greater number of LSTc-SC18 HA contacts, which can be observed in the STD NMR measurements and higher measured binding affinity when compared to LSTc-NY18 HA. The constraints on the avian receptor, LSTa, are consistent with the predominant contacts made by HA RBS with Neu5Ac-α(2→3)-Gal motif as measured by STD NMR signals, which permit a higher degree of flexibility to the sugars on the reducing end of this terminal motif (as seen in the ϕ3/ψ3 , ϕ4/ψ4 conformational map in Fig.18).

The results presented here clearly indicate that even one or two amino acid changes in the HA RBS impose different constraints on the conformation and topology of bound glycan receptors, which in turn governs the biochemical

binding specificity and affinity. Based on this evidence, it is important to carefully assess the effects of transferring amino acid changes that lead to specific receptor-binding properties for a given HA to, and from, a completely different strain or subtype. As stated earlier, the receptor specificity of avian- and human-adapted HAs has been broadly classified based solely on a preference for α(2→3) and α(2→6) linked sialic acid. Based on the data presented here, we find that the amino acid composition of the RBS of various avian-adapted HA subtypes would critically govern structural constraints imposed on diverse and distinct sets of glycans expressed in different tissues, consistent with the ability of viruses from H5, H7 and H9 subtypes to infect distinct tissue types. On the other hand, human-adapted HAs share a characteristic binding to glycans expressed in human upper respiratory epithelia (particularly non-ciliated goblet cells) [21][22][28]. This characteristic binding can be explained on the basis of the structural constraints imposed by SC18 and NY18 on LSTc, which, in turn, is reflected by the θ parameter distribution. Given that goblet cells secrete mucins, it is possible that the characteristic binding of human-adapted HAs to these cell types would increase their propensity for aerosol formation and transmission.

Finally, methods and framework presented in this study to measure the restriction imposed by the RBS of different HAs on the conformational space and topology sampled by glycan receptors could be useful tool to enable more exact surveillance of emerging influenza viruses such as H7N9 and H5N1 to closely monitor their ability to bind to human receptors and acquire the capability for human-to-human transmission.

EXPERIMENTAL PROCEDURES

Cloning, baculovirus synthesis, expression and purification of HA. The soluble form of HA was expressed using the Baculovirus Expression Vector System (BEVS). SC18 (A/South Carolina/1/1918; SC18) baculovirus (generated from pAcGP67-SC18-HA plasmid 26, 27) was a gift from Dr. James Stevens. pAcGp67-NY18-HA and pAcGp67-AV18-HA plasmids were generated from pAcGP67-SC18-HA by [Asp225→Gly] and [Asp190→Glu, Asp225→Gly] mutations respectively. Mutagenesis was carried out using QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene). The primers used for mutagenesis were designed using the web based program, PrimerX (http://bioinformatics.org/primerx/), and synthesized by IDT DNA technologies (Coralville, IA). NY18 and AV18 baculoviruses were created from a pAcGP67-NY18-HA and pAcGP67-AV18-HA constructs using Baculogold system (BD Biosciences, San Jose, CA) according to manufacturer's instructions. The baculoviruses were used to infect 300 ml suspension cultures of Sf9 cells (BD Biosciences, San Jose, CA) cultured in BD BaculoGold Max-XP Insect Cell medium (BD Biosciences, San Jose, CA). These cultures were monitored for signs of infection and harvested 4-5 day post-infection. BEVS produces trimeric HA which provides multivalent binding to glycans. The soluble form of HA was purified from the supernatant of infected cells using the protocol described previously [29]. Briefly, the supernatant was concentrated using Centricon Plus-70 centrifugal filters (Millipore, Billerica, MA) and the trimeric HA was recovered from the concentrated cell supernatant using affinity chromatography with columns packed with Ni-NTA beads (Qiagen, Valencia, CA). Eluting fractions that contained HA were pooled and dialyzed overnight

with a 10 mM Tris-HCl, 50 mM NaCl buffer (pH 8.0). Ion exchange chromatography was then performed on the dialyzed samples using a Mono-Q HR10/10 column (GE healthcare, Piscataway, NJ). The fractions containing HA were pooled together and subjected to ultrafiltration using Amicon Ultra 100 K NMWL membrane filters (Millipore, Billerica, MA). The protein was then concentrated and reconstituted in PBS. The purified protein was quantified using Bio-Rad's protein assay (Bio-Rad, Hercules, CA).

NMR Analysis of SC18, NY18, and AV18 with LSTc and LSTa. Saturation Transfer Difference (STD) and tr-NOESY samples were prepared washing the proteins SC18, NY18 and AV18 (1 mg/mL) with buffered solution (150 mM sodium chloride, 100 mM sodium phosphate, 0.3 mM d- EDTA, D2O, pH7.2) using Amicon® Ultra centrifugal filters, 10 KDa membrane (Millipore). Each ligand (LSTc or LSTa) were added to the corresponding protein sample reaching a final molar ratio of 100:1 [glycan receptor-HA] for the STD measurements and of 25:1 [glycan receptor-HA] for the tr-NOESY measurements, the protein concentration for the STD measurements was 0.01 mM while for the trNOESY experiments it was 0.04 mM. NMR spectra were acquired using a Bruker 600 and 900 MHz AVANCE series NMR spectrometer, both equipped with a high sensitivity 5 mm TCI cryoprobe. LSTc and LSTa resonances could be found in fig. 2 and table 1 chapter 2 . For the STD experiments the on-resonance frequency was set at 7.3 ppm [6600 Hz] and the off-resonance frequency at 20.0 ppm [18000 Hz], a train of 40 gaussian-shaped pulses of 50 ms each were applied to produce a selective saturation of 2 s and D1 was 6 s. The number of scans was 1K and the spectral width used was 12626 Hz. The 2D-NOESY experiments were

carried out using 300 ms of mixing time, recorded for quadrature detection in the indirect dimension and acquired using 16 scans per series of 2048x416 data points. The spectra were recorded at 295 K.

Dose dependent direct binding of SC18, NY18 and AV18 to glycan receptors LSTa and LSTc. LS-tetrasaccharide c ((LSTc - Neu5Ac-α(2→6)-Gal-β(1→4)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc)) and LS-tetrasaccharide a (LSTa - Neu5Ac-α(2→3)-Gal-β(1→3)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc; Accurate Chemicals) were biotinylated with EZ-Link Biotin-LC-Hydrazide (Thermo) according to the manufacturer's instructions. Streptavidin-coated High Binding Capacity 384-well plates (Pierce) were loaded to the full capacity of each well by incubating the well with 50 µl of 2.4 µM of biotinylated LSTa or LSTc overnight at 4 °C. Excess glycans were removed through extensive washing with PBS. The trimeric HA unit comprises three HA monomers and the spatial arrangement of the biotinylated glycans in the wells of the streptavidin plate array favours binding to only one of the three HA monomers in the trimeric HA unit. In order to specifically enhance the correct multivalency in the HA-glycan interactions, the recombinant HA proteins were pre-complexed with the primary and secondary antibodies in the molar ratio of 4:2:1 (HA: primary: secondary). The identical arrangement of 4 trimeric HA units in the pre-complex for all the HAs permitted comparison between their glycan binding affinities. A stock solution containing appropriate amounts of Histidine tagged HA protein, primary antibody (Mouse anti 6X His tag IgG) and secondary antibody (HRP conjugated goat anti Mouse IgG (Santacruz Biotechnology) was combined in the ratio 4:2:1 and incubated on ice for 20 min. Appropriate amounts of pre-complexed stock HA were

diluted to 250 µl with 1% BSA in PBS. 50 µl of this pre-complexed HA was added to each of the glycan-coated wells and incubated at room temperature for 2 hrs followed by the wash steps described above. The binding signal was determined based on HRP activity using Amplex Red Peroxidase Assay (Invitrogen, CA) according to the manufacturer's instructions. The experiments were conducted in triplicate. Minimal binding signals were observed in the negative controls including binding of pre-complexed unit to wells without glycans and binding of the antibodies alone to the wells with glycans. The data obtained from this analysis can be found in Fig. 4.

Molecular Dynamic simulations. The dynamic and conformational properties of the interaction between LSTc and HA was studied by comparing the MD simulation trajectory of three complexes: LSTc-SC18, LSTc-NY18 and LSTc-AV18. The X-ray co-crystal structures of SC18 and NY18 were recently solved with LSTa and LSTc [30]. However, when we started our work, the available co-crystal structures were that of SC18 with LSTc (PDB ID: 2WRG), a swine H1N1 HA (A/swine/Iowa/30) with LSTa and LSTc, 45 ma These X-ray co-crystal structures constituted reasonable starting models of HA-glycan complexes for MD simulations. The LSTc-SC18 complex was built starting with LSTc co-crystallized with SC18 HA (2WRG; coordinates were available for a tetrasaccharide Neu5Ac- α(2→6)-Gal-β(1→4)- GlcNAc- β(1→3)-Gal-) [31]. The protein, HA portion, of the complex was taken directly from the pdb coordinates, selecting a sequence between 60 to 260 amino acid (2WRG numbering) that includes the HA RBS. The solution conformation of LSTc, determined previously in the first part of our work, was superposed onto the

co- crystallized glycan structure, with the non-reducing residues - Neu5Ac and Gal-1 giving a RMSD = 6.5Å. The remaining complexes were obtained from the previously built LSTc-SC18 applying the single mutation Asp225-Gly in silico to generate LSTc-NY18 and the double mutation Asp225-Gly, Asp190-Glu to obtain LSTc-AV18. Is important to note that at the start of the MD simulation, the three complexes are characterized by exactly the same conformation of LSTc and HA with the exception of mutated residues.

LSTa-AV18 was built superimposing a previously selected conformation of LSTa 28 on the LSTa like trisaccharides co-crystallized with H1 (A/Puerto Rico/8/34 H1N1) with amino acids within the binding site typical of an avian HA (pdb code 1RVX). The RMSD between the superposed glycans is 6.2Å calculated on the Neu5Ac and Gal-1 residues. The previously built AV18 protein was superposed to the HA in 1RVX structure matching the protein Cα backbone (RMSD = 0.38Å); then the complex was built taking LSTa and AV18. LSTa/NY18 complex was built from the latter (LSTa/AV18) substituting Glu190 with Asp. Even in that case the model complexes LSTa/AV18 and LSTa/NY18 have the same geometry at the beginning of the MD simulations.

Ambertools 1.4 was used to build the force field. GLYCAM06/Amber was used to describe the glycan and protein part of the complexes, respectively. The simulation cell was built enveloping each macromolecule by a water layer (TIP3P water model) 15 Å wide in three directions, resulting in an orthogonal cell with an edge of approximately 100 Å. Non-bonded potential energy was described using the standard cut-off (12 Å) technique for both electrostatic and dispersive interactions. Each cell was minimized using 100 K steps of the default minimization algorithm included in NAMD. Then 1 ns of MD simulation sampling the NPT ensemble, was used to equilibrate the cell density. The

simulation temperature was set at T = 295K and maintained by a Langevin thermostat as implemented in NAMD, while the Nosé-Hoover Langevin piston algorithm controlled the pressure (1.01325 bar) applied on the cell walls. During the minimization and the cell density equilibration steps, an harmonic potential energy restrain (harmonic constant 50 Kcal/mol) on all the atoms of the complex was applied, while the water molecules were allowed to move freely. The MD simulation for all the modelled complexes was about 120 ns and was completed by applying a soft harmonic restrain on the HA backbone atoms (Cα, N, carbonyl carbon) with a harmonic constant of 2.0 Kcal/mol. This allows the ligand and the side chain residues to be adjusted, while the secondary structure elements are maintained. The MD simulation trajectory was sampled every 10 ps and the comparison between the different complexes were done monitoring selected distances between the ligand and the HA active site residues.

MD simulation trajectory principal component analysis (PCA). PCA was used with the aim of extracting distinct and independent motional modes, it has been used in our study to identify the final bound states in the LSTx-HA complexes. Each frame of the glycan-HA md trajectories were converted from Cartesian coordinates to a distance matrix, measured between the glycan and protein. The distances were determined between the non carbon and hydrogen atoms of the glycan (excluding the glycosidic linkage oxygens and including the carbon of the N-acetyl groups) and the amino acid side chains of HA, the last carbon in the amino acid side chains. A 6 Å cut-off was applied to the distance matrices, this means that only glycan-HA interactions were observed and not the glycan rearranging in solution, away from the RBS. These matrices were

then converted in to a vector and all of them, for a single MD-trajectory, were placed in to a matrix, this dataset was then mean-centered before principal component analysis could be performed. Density based cluster analysis was performed on the first two component loadings (the most significant), the time each cluster appeared in the MD trajectory could then be compared. The first 20 ns of each trajectory was discarded, this left the time interval 20 to 120 ns to be investigated, which was decomposed in to 10000 distance matrices that were examined by PCA. This approach allows the evolution of the glycan-HA complex to be observed, from the initial state to the final state. For the final bound states the ϕ, ψ and θ angles were determined as well as the average glycan-protein contacts for that subset of conformers. The glycan-contacts are represented as networks, with the edge thickness being proportional to distance, the thicker the edge the closer the vertices are, these can be found in Fig. 11, 13 and 14.

Parameters that define glycan conformation and topology. The torsional angles (phi (ϕ) and psi (ψ)) are defined as the following pairs of dihedral angles: $\phi1/\psi1$, $\phi2/\psi2$, $\phi3/\psi3$ and $\phi4/\psi4$, starting from the non-reducing termini. For LSTa the first pair is defined as $\phi1$ = C1-C2-O3-C3, and $\psi1$ = C2-O3-C3- H3, while for LSTc $\phi1$ = C1-C2-O6-C6 and $\psi1$ = C2-O6-C6-C5, as previously defined by Xu et al. [32]. Thereafter successive pairs are defined as $\phi i/\psi i$ = H1-C1-O4'-C'4/C1-O4'-C'4-H4' for the (1→4 linkage) or H1-C1-O3'-C'3/C1-O3'-C'3-H3' for the (1→3 linkage). Atoms labelled by an apostrophe belong to the monosaccharide on the reducing side of the glycosidic linkage, while atoms

without are on the non-reducing side of the glycosidic linkage. The torsional angles are for the LSTc-HA and LSTa-HA complexes are reported in Tables 1 and 2, respectively. The angles were determined by cluster analysis of the data illustrated in Fig. 17 and 18, the analysis used nonparametric density estimation [33] to determine the members of each cluster. The angles represented in the tables are the average values for each ϕ and ψ cluster.

The topology of LSTa and LSTc is defined using a θ angle parameter. The θ angle is defined by the C2, C1 and C1 atoms of the three residues Neu5Ac, Gal and GlcNAc (N-acetyl glucosamine), going from the non-reducing to the reducing end (Figure 1).

REFERENCES

[1] S. Elli, E. Macchi, T. R. Rudd, R. Raman, G. Sassaki, K. Viswanathan, E. a Yates, Z. Shriver, A. Naggi, G. Torri, R. Sasisekharan, and M. Guerrini, "Insights into the human glycan receptor conformation of 1918 pandemic hemagglutinin-glycan complexes derived from nuclear magnetic resonance and molecular dynamics studies.," *Biochemistry*, vol. 53, no. 25, pp. 4122–35, Jul. 2014.

[2] R. Ahmed, M. B. a Oldstone, and P. Palese, "Protective immunity and susceptibility to infectious diseases: lessons from the 1918 influenza pandemic.," *Nat. Immunol.*, vol. 8, no. 11, pp. 1188–93, Nov. 2007.

[3] R. Pérez Velasco, N. Praditsitthikorn, K. Wichmann, A. Mohara, S. Kotirum, S. Tantivess, C. Vallenas, H. Harmanci, and Y. Teerawattananon, "Systematic review of economic evaluations of preparedness strategies and interventions against influenza pandemics.," *PLoS One*, vol. 7, no. 2, p. e30333, Jan. 2012.

[4] C. Fraser, C. A. Donnelly, S. Cauchemez, W. P. Hanage, D. Van Kerkhove, T. D. Hollingsworth, J. Griffin, R. F. Baggaley, E. Jenkins, E. J. Lyons, T. Jombart, W. R. Hinsley, N. C. Grassly, F. Balloux, A. C. Ghani, N. M. Ferguson, A. Rambaut, G. Oliver, H. Lopez-gatell, C. M. Alpuche-aranda, I. B. Chapela, P. Zavala, D. Ma, E. Guevara, F. Checchi, and

E. Garcia, “Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings,” *Science*, vol. 324, no. 5934, pp. 1557–1561, 2013.

[5] Y. Itoh, K. Shinya, M. Kiso, T. Watanabe, Y. Sakoda, M. Hatta, Y. Muramoto, D. Tamura, Y. Sakai-Tagawa, T. Noda, S. Sakabe, M. Imai, Y. Hatta, S. Watanabe, C. Li, S. Yamada, K. Fujii, S. Murakami, H. Imai, S. Kakugawa, M. Ito, R. Takano, K. Iwatsuki-Horimoto, M. Shimojima, T. Horimoto, H. Goto, K. Takahashi, A. Makino, H. Ishigaki, M. Nakayama, M. Okamatsu, K. Takahashi, D. Warshauer, P. a Shult, R. Saito, H. Suzuki, Y. Furuta, M. Yamashita, K. Mitamura, K. Nakano, M. Nakamura, R. Brockman-Schneider, H. Mitamura, M. Yamazaki, N. Sugaya, M. Suresh, M. Ozawa, G. Neumann, J. Gern, H. Kida, K. Ogasawara, and Y. Kawaoka, “In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses.,” *Nature*, vol. 460, no. 7258, pp. 1021–5, Aug. 2009.

[6] M. B. Pearce, A. Jayaraman, C. Pappas, J. A. Belser, H. Zeng, and K. M. Gustin, “Pathogenesis and transmission of swine origin A (H3N2) v in fl uenza viruses in ferrets,” *Proc. Natl. Acad. Sci.*, vol. 109, no. 10, pp. 3944–3949, 2012.

[7] Y. Chen, W. Liang, S. Yang, N. Wu, H. Gao, J. Sheng, H. Yao, J. Wo, Q. Fang, D. Cui, Y. Li, X. Yao, Y. Zhang, H. Wu, S. Zheng, H. Diao, S. Xia, Y. Zhang, K.-H. Chan, H.-W. Tsoi, J. L.-L. Teng, W. Song, P. Wang, S.-Y. Lau, M. Zheng, J. F.-W. Chan, K. K.-W. To, H. Chen, L. Li, and K.-Y. Yuen, “Human infections with the emerging avian influenza A H7N9 virus from wet market poultry: clinical analysis and characterisation of viral genome.,” *Lancet*, vol. 381, no. 9881, pp. 1916–25, Jun. 2013.

[8] R. a M. Fouchier, P. M. Schneeberger, F. W. Rozendaal, J. M. Broekman, S. a G. Kemink, V. Munster, T. Kuiken, G. F. Rimmelzwaan, M. Schutten, G. J. J. Van Doornum, G. Koch, A. Bosman, M. Koopmans, and A. D. M. E. Osterhaus, “Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 5, pp. 1356–61, Feb. 2004.

[9] R. Gao, B. Cao, Y. Hu, Z. Feng, D. Wang, W. Hu, J. Chen, Z. Jie, H. Qiu, K. Xu, X. Xu, H. Lu, W. Zhu, Z. Gao, N. Xiang, Y. Shen, Z. He, Y. Gu, Z. Zhang, Y. Yang, X. Zhao, L. Zhou, X. Li, S. Zou, Y. Zhang, X. Li, L. Yang, J. Guo, J. Dong, Q. Li, L. Dong, Y. Zhu, T. Bai, S. Wang, P. Hao, W. Yang, Y. Zhang, J. Han, H. Yu, D. Li, G. F. Gao, G. Wu, Y. Wang, Z. Yuan, and Y. Shu, “Human infection with a novel avian-origin influenza A (H7N9) virus.,” *N. Engl. J. Med.*, vol. 368, no. 20, pp. 1888–97, May 2013.

[10] K. Subbarao* and J. Katz, “Avian influenza viruses infecting humans,” *Cell. Mol. Life Sci.*, vol. 57, no. 12, pp. 1770–1784, Nov. 2000.

[11] H. Wan, E. M. Sorrell, H. Song, M. J. Hossain, G. Ramirez-Nieto, I. Monne, J. Stevens, G. Cattoli, I. Capua, L.-M. Chen, R. O. Donis, J. Busch, J. C. Paulson, C. Brockwell, R. Webby, J. Blanco, M. Q. Al-Natour, and D. R. Perez, “Replication and transmission of H9N2 influenza viruses in ferrets: evaluation of pandemic potential.,” *PLoS One*, vol. 3, no. 8, p. e2923, Jan. 2008.

[12] T. Watanabe, M. Kiso, S. Fukuyama, N. Nakajima, M. Imai, S. Yamada, S. Murakami, S. Yamayoshi, K. Iwatsuki-Horimoto, Y. Sakoda, E. Takashita, R. McBride, T. Noda, M. Hatta, H. Imai, D. Zhao, N. Kishida, M. Shirakura, R. P. de Vries, S. Shichinohe, M. Okamatsu, T. Tamura, Y. Tomita, N. Fujimoto, K. Goto, H. Katsura, E. Kawakami, I. Ishikawa, S. Watanabe, M. Ito, Y. Sakai-Tagawa, Y. Sugita, R. Uraki, R. Yamaji, A. J. Eisfeld, G. Zhong, S. Fan, J. Ping, E. a Maher, A. Hanson, Y. Uchida, T. Saito, M. Ozawa, G. Neumann, H. Kida, T. Odagiri, J. C. Paulson, H. Hasegawa, M. Tashiro, and Y. Kawaoka, "Characterization of H7N9 influenza A viruses isolated from humans.," *Nature*, vol. 501, no. 7468, pp. 551–5, Sep. 2013.

[13] J. a Belser, O. Blixt, L.-M. Chen, C. Pappas, T. R. Maines, N. Van Hoeven, R. Donis, J. Busch, R. McBride, J. C. Paulson, J. M. Katz, and T. M. Tumpey, "Contemporary North American influenza H7 viruses possess human receptor specificity: Implications for virus transmissibility.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 21, pp. 7558–63, May 2008.

[14] T. R. Maines, L. Chen, Y. Matsuoka, H. Chen, T. Rowe, J. Ortin, A. Falco, N. Tran, L. Quynh, E. R. Sedyaningsih, S. Harun, T. M. Tumpey, R. O. Donis, N. J. Cox, K. Subbarao, and J. M. Katz, "Lack of transmission of H5N1 avian – human reassortant influenza viruses in a ferret model," *Proc. Natl. Acad. Sci.*, vol. 103, no. 32, pp. 12121–12126, 2006.

[15] A. S. Gambaryan, A. B. Tuzikov, V. E. Piskarev, S. S. Yamnikova, and D. K. Lvov, "Specification of Receptor-Binding Phenotypes of Influenza Virus Isolates from Different Hosts Using Synthetic Sialylglycopolymers : Non-Egg-Adapted Human H1 and H3 Influenza A and Influenza B Viruses Share a Common High Binding Affinity for 6 -Sialyl (," vol. 350, no. 232, pp. 345–350, Jun. 1997.

[16] R. J. Russell, D. J. Stevens, L. F. Haire, S. J. Gamblin, and J. J. Skehel, "Avian and human receptor binding by hemagglutinins of influenza A viruses.," *Glycoconj. J.*, vol. 23, no. 1–2, pp. 85–92, Feb. 2006.

[17] K. Shinya, M. Ebina, S. Yamada, M. Ono, N. Kasai, and Y. Kawaoka, "Avian flu: influenza virus receptors in the human airway.," *Nature*, vol. 440, no. 7083, pp. 435–6, Mar. 2006.

[18] J. Stevens, O. Blixt, L. Glaser, J. K. Taubenberger, P. Palese, J. C. Paulson, and I. a Wilson, "Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities.," *J. Mol. Biol.*, vol. 355, no. 5, pp. 1143–55, Feb. 2006.

[19] A. Chandrasekaran, A. Srinivasan, R. Raman, K. Viswanathan, S. Raguram, T. M. Tumpey, V. Sasisekharan, and R. Sasisekharan, "Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin.," *Nat. Biotechnol.*, vol. 26, no. 1, pp. 107–13, Jan. 2008.

[20] Z. Shriver, R. Raman, K. Viswanathan, and R. Sasisekharan, “Context-specific target definition in influenza a virus hemagglutinin-glycan receptor interactions.,” *Chem. Biol.*, vol. 16, no. 8, pp. 803–14, Aug. 2009.

[21] A. Srinivasan, K. Viswanathan, R. Raman, A. Chandrasekaran, S. Raguram, T. M. Tumpey, V. Sasisekharan, and R. Sasisekharan, “Quantitative biochemical rationale for differences in transmissibility of 1918 pandemic influenza A viruses.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 8, pp. 2800–5, Feb. 2008.

[22] A. Jayaraman, A. Chandrasekaran, K. Viswanathan, R. Raman, J. G. Fox, and R. Sasisekharan, “Decoding the distribution of glycan receptors for human-adapted influenza A viruses in ferret respiratory tract.,” *PLoS One*, vol. 7, no. 2, p. e27517, Jan. 2012.

[23] C. A. Bewley, “Illuminating the switch in influenza viruses,” *Nat. Biotechnol.*, vol. 26, pp. 60–62, Dec. 2008.

[24] T. M. Tumpey, T. R. Maines, N. Van Hoeven, L. Glaser, A. Solórzano, C. Pappas, N. J. Cox, D. E. Swayne, P. Palese, J. M. Katz, and A. García-Sastre, “A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission.,” *Science*, vol. 315, no. 5812, pp. 655–9, Feb. 2007.

[25] I. Seema S. Lakdawalaa, Angela R. Shiha, Akila Jayaramanb, Elaine W. Lamirandea and and K. S. Moorea, Myeisha Paskela, Ram Sasisekharanb, “Receptor Specificity does not affect Replication or Virulence of the 2009 Pandemic H1N1 Influenza Virus in Mice and Ferrets.pdf,” *Virology*, vol. 446, pp. 349–356, 2013.

[26] F. J. C. and J. J. B. Virginia Roldòs, “Carbohydrate–Protein Interactions- A 3D View by NMR.pdf,” *ChemBioChem*, vol. 12, pp. 990–1005, 2011.

[27] G. L. Sassaki, S. Elli, T. R. Rudd, E. Macchi, E. A. Yates, A. Naggi, Z. Shriver, R. Raman, R. Sasisekharan, G. Torri, and M. Guerrini, “Human (alfa2->6) and avian (alfa2->3) sialylated receptors of influenza A virus show distinct conformations and dynamics in solution,” *Biochemistry*, vol. 52, no. 41, pp. 7217–7230, 2013.

[28] D. van Riel, V. J. Munster, E. de Wit, G. F. Rimmelzwaan, R. a M. Fouchier, A. D. M. E. Osterhaus, and T. Kuiken, “Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals.,” *Am. J. Pathol.*, vol. 171, no. 4, pp. 1215–23, Oct. 2007.

[29] J. Stevens, A. L. Corper, C. F. Basler, J. K. Taubenberger, P. Palese, and I. A. Wilson, “Structure of the Uncleaved Human H1 Hemagglutinin from the Extinct 1918 Influenza Virus,” *Science*, vol. 303, no. MARCH, pp. 1866–1871, 2004.

[30] W. Zhang, Y. Shi, J. Qi, F. Gao, Q. Li, Z. Fan, J. Yan, and G. F. Gao, “Molecular basis of the receptor binding specificity switch of the hemagglutinins from both the 1918 and 2009 pandemic influenza A viruses by a D225G substitution.,” *J. Virol.*, vol. 87, no. 10, pp. 5949–58, May 2013.

[31] J. Liu, D. J. Stevens, L. F. Haire, P. a Walker, P. J. Coombs, R. J. Russell, S. J. Gamblin, and J. J. Skehel, “Structures of receptor complexes formed by hemagglutinins from the Asian Influenza pandemic of 1957.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 40, pp. 17175–80, Oct. 2009.

[32] D. Xu, E. I. Newhouse, R. E. Amaro, H. C. Pao, L. S. Cheng, P. R. L. Markwick, J. A. McCammon, W. W. Li, and P. W. Arzberger, “Distinct glycan topology for avian and human sialopentasaccharide receptor analogues upon binding different hemagglutinins: a molecular dynamics perspective.,” *J. Mol. Biol.*, vol. 387, no. 2, pp. 465–91, Mar. 2009.

[33] A. Azzalini and G. Menardi, “Clustering via Nonparametric Density Estimation :,” *J. Stat. Softw.*, vol. 57, no. 11, 2014.

NMR and MD investigation of a new avian virus strain H7N9 and its interaction with avian and human receptor

INTRODUCTION

In February 2013, a new influenza A virus subtype H7N9 infected humans mainly in China, resulting in rapidly progressing severe respiratory infection and high mortality. Infection of poultry with influenza A subtype H7 virus occurs worldwide, but the introduction of this subtype to humans in Asia has not been observed previously[1].

Even though it was classified as a low pathogenicity avian influenza, the virus was able to rapidly infect more than 100 people, 40 of which died and many were severely ill. Infection seems to involve people being in contact with infected poultry, a direct human-to-human transmission was never observed, recently evidence was found of a virus isolated from humans that was able to transmit efficiently between ferrets by respiratory droplet [2]. Others problems were due to the nonpathogenic nature of the virus (H7N9) in poultry, which enables the avian H7N9 virus to replicate 'silently' in avian species and than transmit to humans. Its replication in humans provides further opportunities for the virus to acquire more mutations, becoming more virulent and transmissible in the human population. All these reasons explain the concern regarding the transmissibility of the virus between humans. Two different

viruses were isolated in human: A/Shanghai/1/2013 (SH-H7N9), characterized by the "avian signature" Gln^{226}, and the A/Anhui/1/2013 (AH-H7N9) characterized by the "human signature" Leu^{226}.

Gene	Sites	Position	A/Shanghai/1/2013	A/Anhui/1/2013
HA	Cleavege site RBS positions (H3 numbering), altered receptor specificity Q226L G228S Glycosylation motifs	 226 228 30NGTK, 46NATE, 249NDTV, 421NWTR, 493NNTY (conserved in H7 HA viruses)	PEIPKGR*G Q G 30NGTK, 46NATE, 249NDTV, 421NWTR, 493NNTY (conserved in H7 HA viruses)	PEIPKGR*G L G 30NGTK, 46NATE, 249NDTV, 421NWTR, 493NNTY (conserved in H7 HA viruses)

Table 1. Molecular analysis of two of the 2013 H7N9 Viruses, reproduced from Gao et al. [1]

It has been demonstrated [3] that the Shanghai and Anhui H7N9 viruses were able to bind both avian and human receptors, with the Shanghai virus having a strong preference for the avian receptor, whereas, Anhui shows a comparable affinity for both avian and human receptors [4][4].

It is interesting to note that the H7N9 HA subtypes exhibit a smaller affinity toward avian and human glycan receptors compared to H1N1 and H5N1.

It is also true that all these data could describe the attempt of the avian A virus H7N9 to progressively acquire, by amino acid mutation, the affinity for the human receptor, until now this has been achieved without losing affinity towards the avian receptor, a behavior not common to potentially pandemic

virus in human host, such as H1N1 (Fig. 1), that loose affinity toward the avian receptor. Other studies [5] reported an increased in human receptor binding due to a single Gly228-Ser amino acid change.

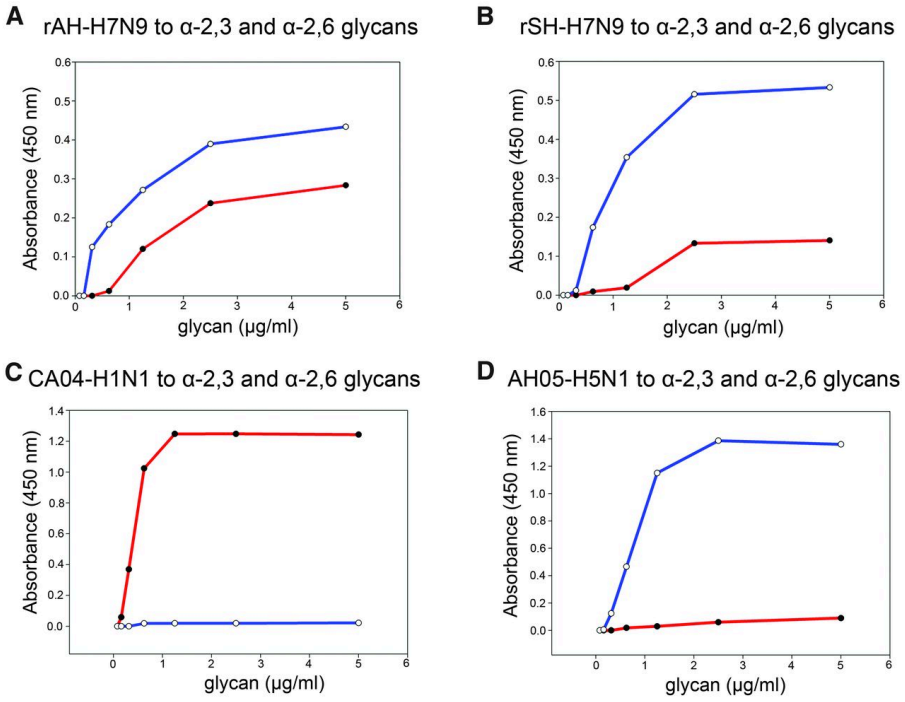


Fig. 1 Receptor binding properties at virus level. Binding of virus to α-2,3-linked or α-2,6-linked receptor was determinated by solid phase binding assay. (A) rAH-H7N9(reverse genetics-rescued A/Anhui/1/2013) virus (B) rSH-H7N9 (reverse genetics-rescued A/Shanghai/1/2013) virus.(C) CA04-H1N1 virus; (D) AH05-H5N1 virus. Blu, binding to α-2,3, red binding to α-2,6 [4]. Reproduced from Shi et.al[4]

The glycine in position 228 of AH-H7N9 does not possess the inter-amino acid network, which instead seems to be important in another virus H3 HA (A/Aichi/1/1968, pandemic virus) that has a serine in position 228. This Ser228 in H3 of Aichi68 is required for the inter-amino acid hydrogen bonding network

involving Ser186, Thr187 and Glu190, which positions Glu190 to make critical contacts with the Neu5Ac.

The substitution Gly186-Val, also increases the hydrophobicity of the binding site, consequently favoring interactions with human receptors [3].

Even thought a lot of work has been undertaken, regarding the H7N9 subtype, the molecular base by which H7 switches preferences between human (α2-6) and avian (α2-3) glycan receptor, remains poorly understood; even though the evolution tendency of the H7N9 virus is to extend its infection ability.

We combined NMR and molecular dynamic simulation in order to investigate the solution structure of LSTc and LSTa bound to H7N9wt with the aim to understand the H7 molecular recognition ability toward LSTa and/or LSTc. H7N9 binding ability will be compared on its previously cited mutants. H7N9sm (Gly228-Ser) and H7N9dm (Gly228-Ser and Val186-Gly) in order to understand their effects on the H7 glycan recognition ability. Finally, a comparison between H7N9wt and H1N1(SC18) RBS and their ability to binding to the avian and human glycan receptors will be made.

RESULTS AND DISCUSSION

NMR analysis of HA-glycan interactions. NMR spectra were obtained using a 900 MHz spectrometer, equipped with a 5mm cryoprobe.

Saturation-transfer difference (STD) experiments were performed in order to obtain a description of the epitope mapping of the interaction between the human (LSTc) and avian (LSTa) receptors and HAs. The NMR analysis of the LSTc and LSTa in complex with H7N9 wild type (Fig. 2-4) indicated that both the receptors interact with HA, principally through the non-reducing terminal sialic

acid (Neu5Ac). Nevertheless, a deeper analysis of the STD spectra, revealed some differences in the epitope mapping of the interaction between LSTc and LSTa interacting with H7N9wt.

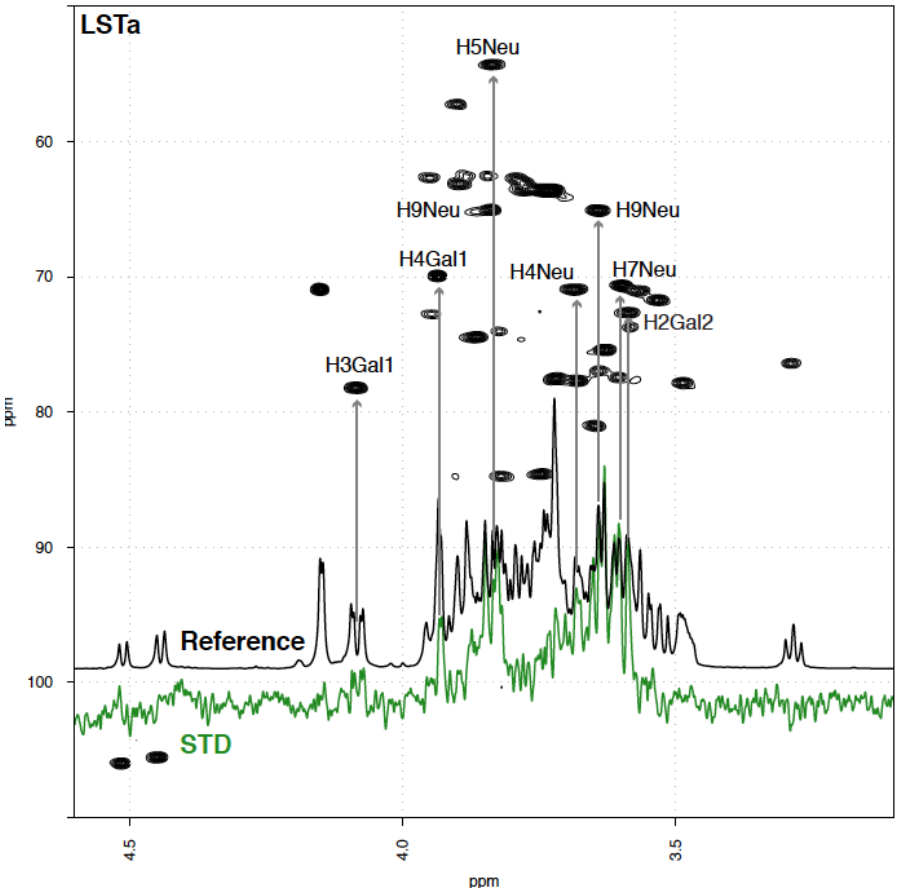


Fig.2 Main region of the overlaid STD HSQC spectra of LSTa receptor complex. STD spectrum of LSTa-H7N9wt (green), while the correspondent reference spectrum is in black, superimpose to the HSQC spectrum of LSTa.

The ^{1}H STD spectrum of LSTa bound to H7N9wt was superimposed on the HSQC spectrum of LSTa and to the correspondent reference spectrum of STD (Fig. 2

and Fig. 4 left) in order to better assign the STD signals. The same was reported for LSTc (Fig. 3 and Fig. 4 right).

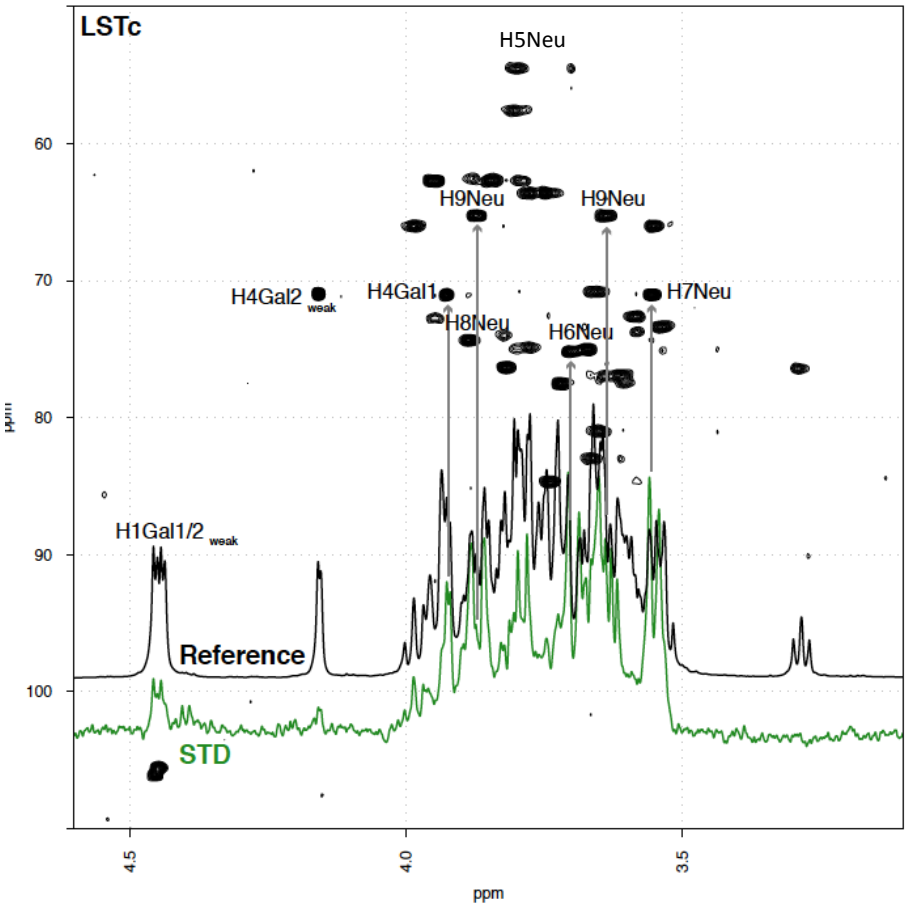


Fig.3 Main region of the overlaid STD HSQC spectra of LSTc receptor complex. STD spectrum of LSTc-H7N9wt (green), while the correspondent reference spectrum is in black, superimpose to the HSQC spectrum of LSTc.

In addition to the non-reducing end Neu5Ac, LSTa interacts with H7N9 also through H3 and H4 of Gal1 and H2 of Gal2 (Fig. 2), unfortunately due to the overlap of the signals of the CH_3 of Neu5Ac and the CH_3 of GlcNAc, it is not

possible to distinguish which signal is more involved in the interaction (Fig. 4 left).

While, LSTc bound H7N9wt via H1 and H4 of Gal1 and Gal2, even if these signals are weaker compared to those signals belonging to the Neu5Ac (Fig. 3).

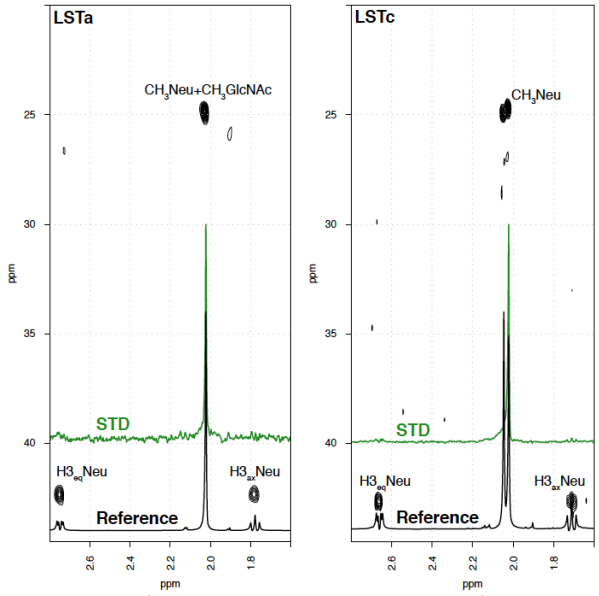


Fig.4 N-Acetyl regions of the overlaid STD HSQC spectra of LSTa and LSTc- receptor complex with H7N9wt. STD spectrum of LSTa-H7N9wt (left) and of LSTc-H7N9wt (right) are in green overlaid upon the HSQC spectrum of LSTa (left) and LSTc (right). In black the correspondent reference spectra.

Particularly, in this case it is possible to distinguish between the two methyl groups (Fig. 4 right), and only the methyl group of Neu5Ac appears in the STD spectrum of LSTc-H7N9wt complex, indicating the proximity of Neu5Ac methyl group to the protein; whiles no signal belonging to the methyl group of GlcNac appears in the STD spectrum.

The NMR glycan-HA interaction study was performed involving also the H7N9 mutants. The tested complexes include: AH-H7N9sm (Gly228-Ser, single

mutant) and AH-H7N9dm (Gly228-Ser and Val186-Gly, double mutant) with human (LSTc) and avian (LSTa) receptors.

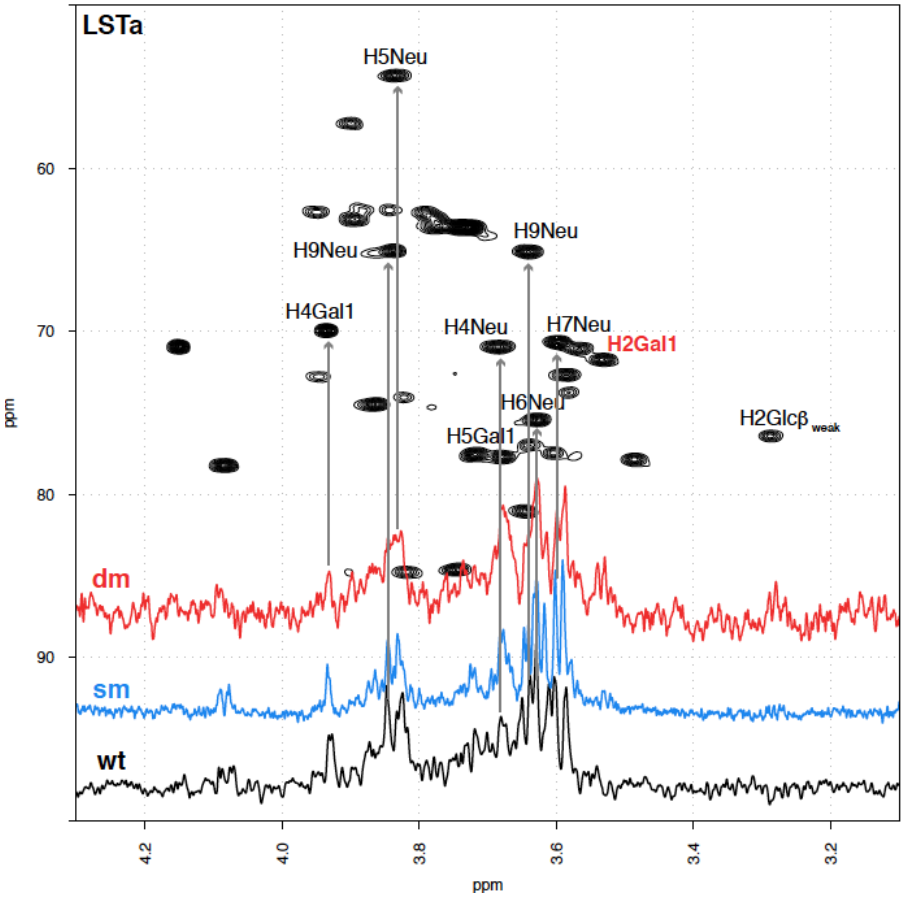


Fig. 5 In black the STD spectrum of LSTa/H7N9wt complex, in blue the STD spectrum of LSTa/H7N9sm and in red the STD spectrum of LSTa/H7N9dm superimposed to the HSQC spectrum (black) of LSTa.

The comparison of the STD spectra of the complexes between wt, sm and dm and LSTa are reported in Fig. 5 and 7 (left) for LSTa, whereas those between wt, sm and dm and LSTc are reported in Fig. 6 and 7 (right).

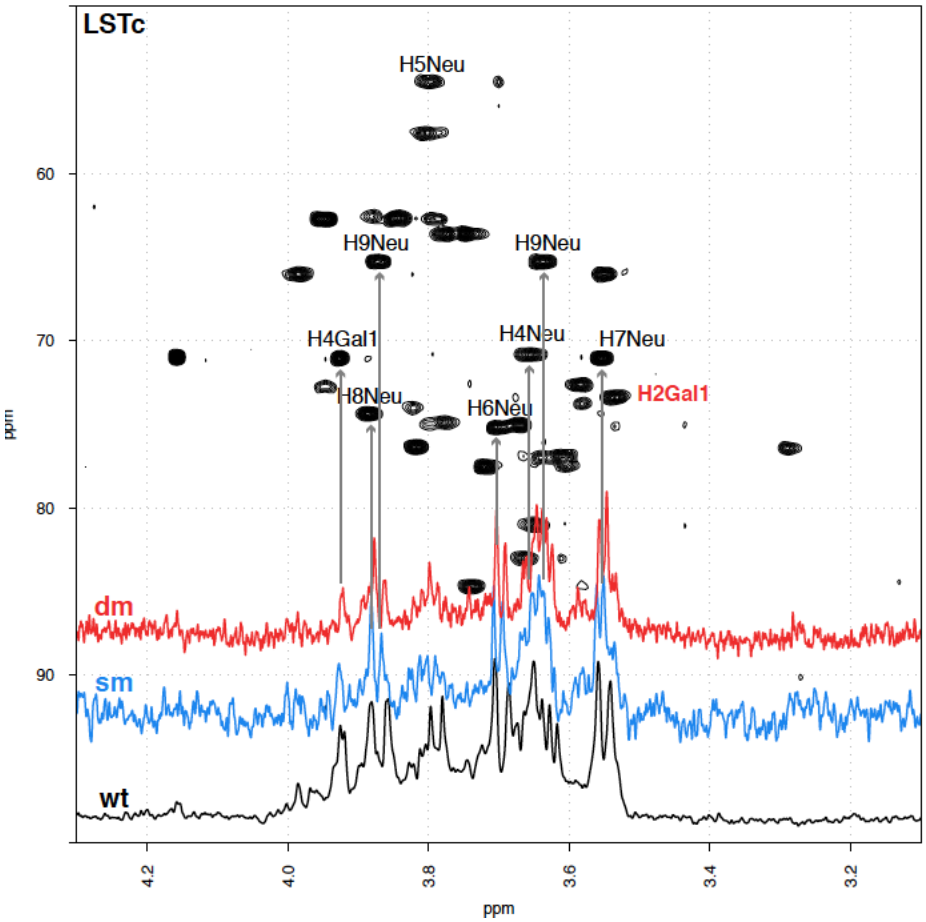


Fig. 6 In black the STD spectrum of LSTc /H7N9wt complex, in blue the STD spectrum of LSTc/H7N9sm and in red the STD spectrum of LSTc/H7N9dm superimposed to the HSQC spectrum (black) of LSTc.

The STD spectra of the LSTa-H7N9sm and LSTa-H7N9wt complexes show a similar profile (Fig. 5, blue and black lines), indicating that the sugar residue involved in the interaction of the sm and the wt with LSTa is principally the Neu5Ac. On the other hand, in the STD spectrum of the LSTa-H7N9dm complex

a weak signal belonging to Glc appears (Fig. 5, red line), indicating that the non-reducing end of LSTa could be closer to HA in H7N9dm compared to the wt.

The human receptor, LSTc also interacts with the sm and dm mainly using the Neu5Ac residues, even if also in this case some differences are visible. In the STD spectra of LSTc-H7N9sm and LSTc-H7N9dm complexes (Fig. 6) a signal belonging to H2 of Gal1 appears, while the signal of H5 of Neu5Ac in the sm and dm is weaker compared to the wt. Notably, LSTc interacts with H7N9sm and H7N9dm also with the methyl group of the GlcNAc (Fig. 7 right), even though this interaction is weaker compared to that of the methyl group of the Neu5Ac.

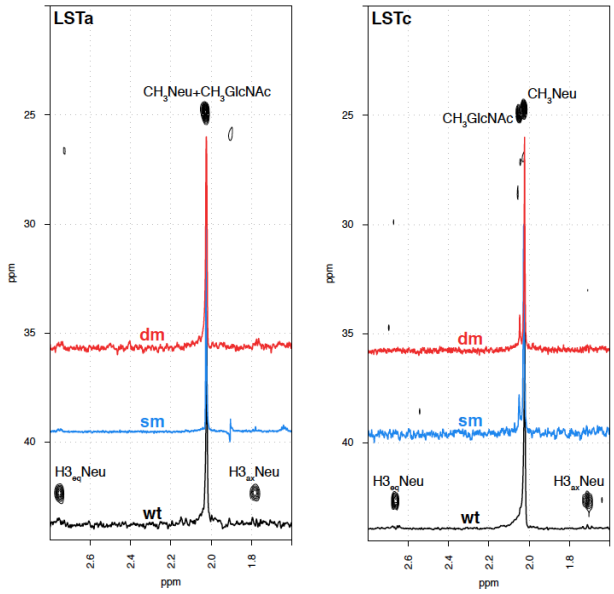


Fig. 7. N-Acetyl regions of the overlaid STD HSQC spectra (black) of LSTa (left) and LSTc (right)-receptor complex with H7N9wt (black), H7N9sm (blue) and H7N9dm (red).

These results indicate a slightly difference in term of interaction between wt, sm and dm and human and avian receptors, respectively.

In order to qualitatively compare H7N9sm and H7N9dm affinity toward LSTc or LSTa, competitive experiments were performed [6] (Fig. 8 and 9). STD-NMR experiments (Fig. 8 and 9 respectively) were recorded on a sample prepared adding to the same NMR tube an equal amount of human and avian glycans to the H7N9sm or H7N9dm.

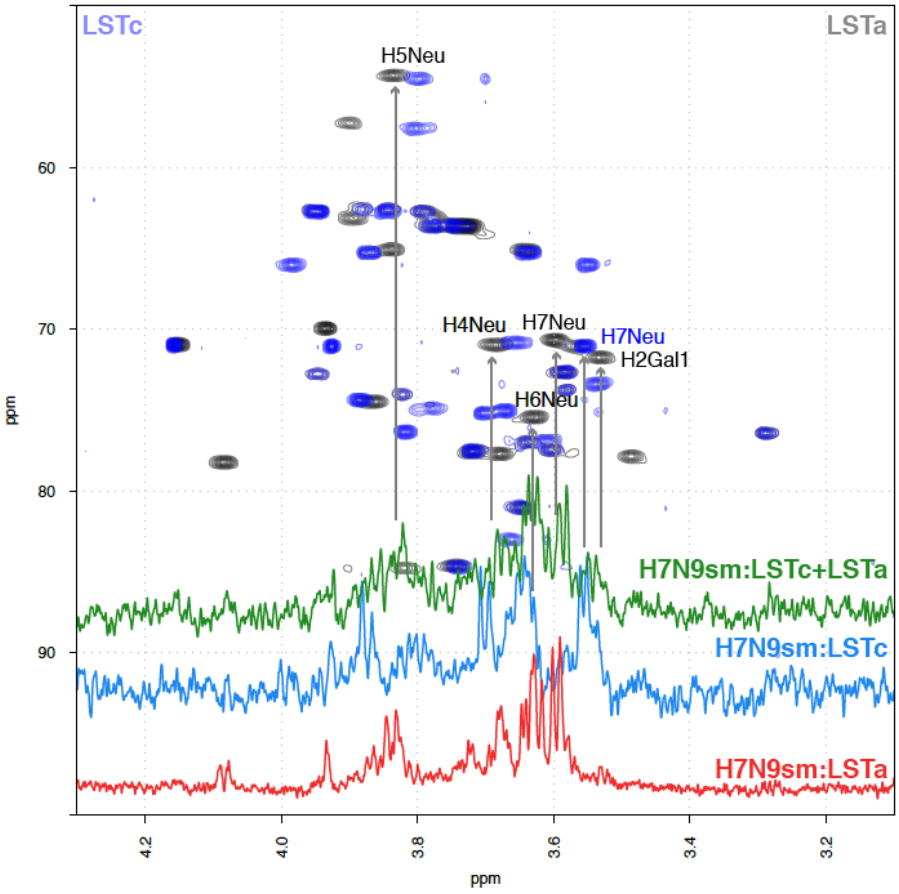


Fig. 8 Competitive experiments. STD NMR experiment between H7N9sm and LSTa (green), H7N9sm and LSTc (purple) and H7N9sm and a mixture of LSTc and LSTa (red) superimposed to HSQC spectra of LSTc (blue) and LSTa (black).

STD signals belonging to LSTa appear slightly stronger to those of LSTc in the spectrum of sample where both glycans are mixed and added to H7N9sm, suggesting that LSTa preferentially binds the single mutant HA (Fig. 8).

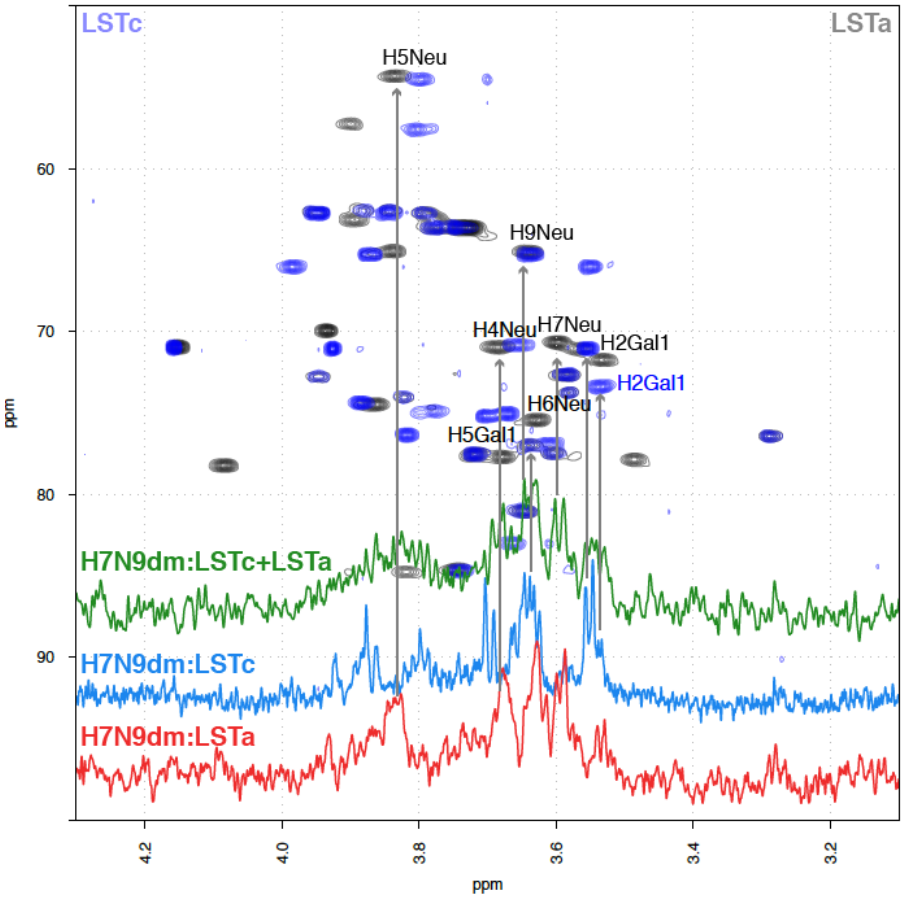


Fig. 9 Competitive experiments. STD NMR experiment between H7N9dm and LSTa (blue), H7N9sm and LSTc (green) and H7N9sm and a mixture of LSTc and LSTa (red) superimposed to HSQC spectra of LSTc (blue) and LSTa (black).

The LSTa and LSTc competitive STD experiment toward the double mutant (H7N9dm) is reported in Fig 9. Although, also in this case the results are difficult

to interpret, due to the low signal-to-noise and many overlapping features, signals belonging to LSTa appear to be prevalent in the STD spectrum.

Titration experiments, largely used to study protein-ligand interaction, are a promising strategy to identify the best ligand in competitive experiments. Even if the lack of sufficient glycans requested for the titration experiments did not allow to perform these part of the study, preliminary data obtained clearly indicate that the studied mutations has not lead H7N9 to acquire preference to bind the human receptor.

Molecular Dynamics Simulation of H7N9wt HA and its mutant forms with Glycan Receptors. This part of the work should be considered preliminary, since during the drafting of the thesis the Molecular Dynamic simulation of LSTc-H7N9dm, LSTa-H7N9sm and LSTa-H7N9dm were still incomplete, 60 ns compared to the 100 ns of the others complexes, even though this time should be sufficient to draw some preliminary conclusions.

MD simulations were used with the aim to optimize the structures of the complexes obtained through X-ray and qualitatively compared the models with the NMR results. The comparison between models of the different complexes (LSTc-H7N9wt, LSTc-H7N9sm, LSTc-H7N9dm, LSTa-H7N9wt, LSTa-H7N9sm, LSTa-H7N9dm) allows to observe how H7N9wt interact with LSTc and LSTa, drawing a 3D structural and dynamic picture of the binding interaction and also how the single (Gly228-Ser) and double mutations (Gly228-Ser, Val186-Gly) modify binding and recognition ability of HAs toward LSTc and LSTa.

It should be noted that the glycan starting geometries are the same for the different MD simulations, corresponding to the solution forms identified previously[7].

Observation of the LSTc-H7N9wt and LSTa-H7N9wt simulations (Fig. 10) indicated that the Neu5Ac is the major residue involved in the interaction in both glycans. In fact in both complexes, this residue maintains, for the whole duration of the MD simulation (approximately 100 ns), its starting position found in the H7 RBS co-crystallized in 4BSE and 4BSF[3].

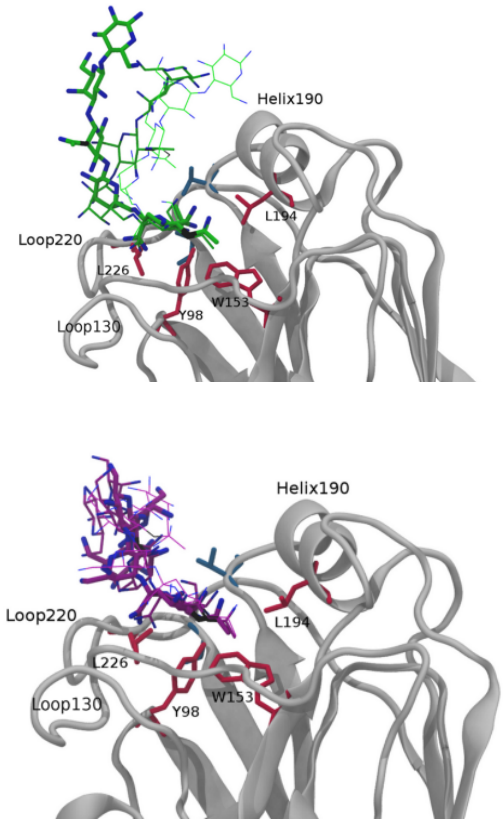


Fig. 10 Snapshot of the LSTc-H7N9wt (top) and LSTa-H7N9wt (bottom) complexes at simulation time of 25ns (small thickness) and 56 ns (medium thickness) and 100 ns (wide tube).

In agreement with literature data [4], the dynamic indicates that the reducing end of LSTc (Gal2-Glc, Fig. 10 in green at the top) shows greater conformational freedom compared to LSTa. The reducing end of LSTa interacts with the groove between loop 220 and helix 190, but the most involved residues remains the Neu5Ac, as observed also in the STD NMR experiments (Fig. 2 and 3).

The differences in LSTc and LSTa glycan binding maps are particularly visible in their reducing ends, where they extend in two different directions: LSTc toward the 190 helix, and LSTa between loop220 and 190 helix (Fig. 10). The hemagglutinin of H7N9wt can be visualized as a hand that catches LSTc or LSTa, while loop130 and loop220 can be imagined as two fingers that cooperate to seize the receptor.

The comparison of the distances between loop 220 and loop130 in LSTa-H7N9wt along the MD simulation (Fig. 11 left), shows smaller fluctuation and average distances in the former compared to the latter, possibly correlated to a slightly stronger interaction. From another point of view the distances between hydrogens of the glycans and H7N9wt in the short range (2 to 11 Å) are slightly more populated in LSTa-H7N9 (Fig. 11 right) indicating a stronger binding interaction of H7N9wt toward the avian receptor.

In the LSTc-H7N9sm complex, the presence of the serine in position 228 instead of glycine, introduces an hydrogen bond at the bottom of the RBS, making the interaction between the Neu5Ac tail (C8 and C9) and the HA stronger, without modifying the Neu5Ac position in H7 RBS, in accordance to the STD experiment (Fig. 6).

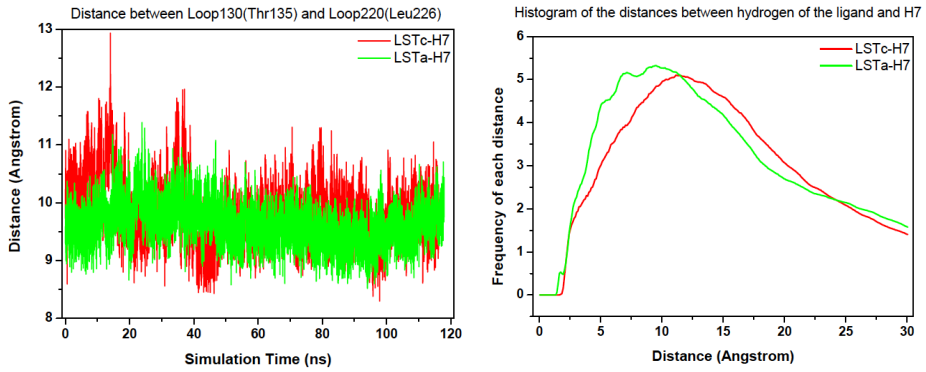


Fig. 11 The h-h histogram is calculated between 50ns until the end of the MD simulation trajectory, using a step = 2 geometries and a selection (protein and hydrogen)

On the contrary the LSTc reducing end modifies its interaction pattern with 190 Helix upon H7 mutation, as can be appreciated in Fig. 12. In this picture the structure of LSTc-H7N9wt and LSTc-H7N9sm complexes at different simulation time are superposed by the protein backbone (ribbon), while the corresponding ligand appears colored in yellow or in blue. LSTc presents a smaller mobility when binds H7N9wt, this suggests that the additional hydrogen bond introduced upon mutation at the RBS bottom perturbs the whole ligand mobility.

In the double mutant the additional mutation at position 186, with glycine replacing a valine, reduces the steric hindrance (hydrophobic surface) between loop220 and helix 190 in the RBS of HA (Fig. 13), allowing LSTc to better adapt its reducing end trisaccharides (GlcNAc-Gal2-Glc) between these secondary structure elements.

The MD simulation description predicts that upon this mutation (G228S and V186G) the distances between LSTc reducing end and H7 RBS surface hydrogens decrease significantly (Fig. 14).

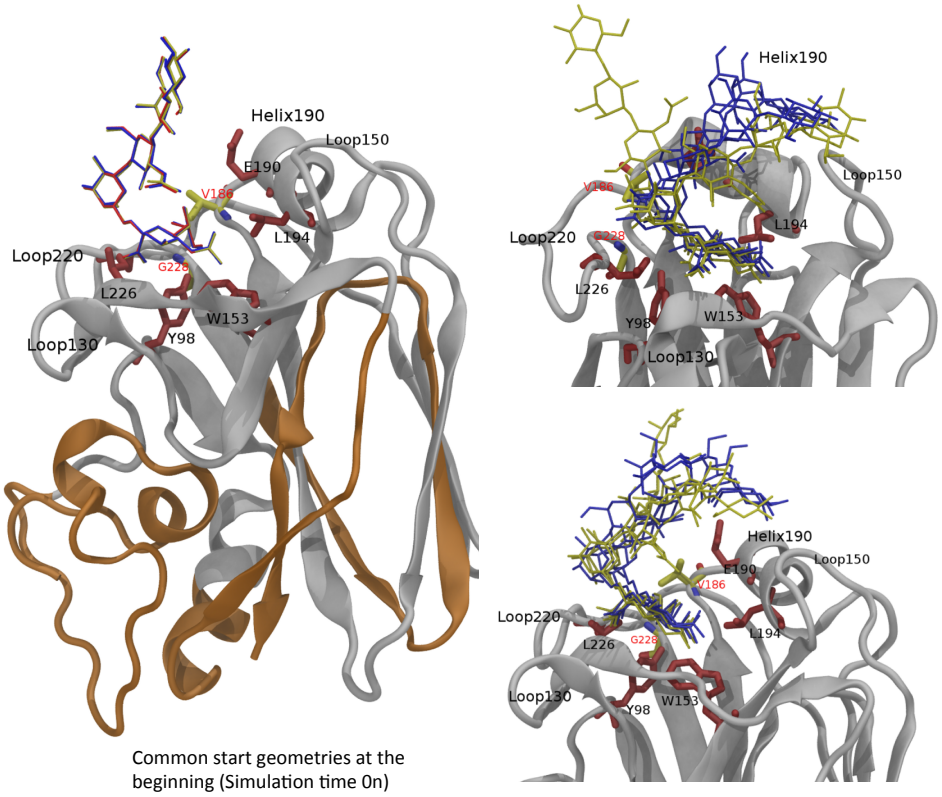


Fig. 12 LSTc-H7N9 complexes (wt and sm) at 41, 51 and 61 ns are superposed by the CA backbone in two different orientations (right). The three different ligand snapshot are shown in yellow (LSTc-H7N9wt) and in blue (LSTc-H7N9sm). The orange ribbon shows the protein backbone restrained by harmonic potential during the MD simulation, while the white ribbon indicate protein backbone not restrained, i.e. free to adjust around the glycan.

The superposition of several snapshots sampled in a comparable timescale, as showed in fig. 12 and 13, give the chance to qualitatively distinguish the two MD trajectories not only in term of conformation, but also from the dynamic behavior. It is important to underline that the two MD simulation trajectories showed in fig. 12 and 13 have at the beginning (T =0 ns) exactly the same complex structure, except the mutation point in the RBS.

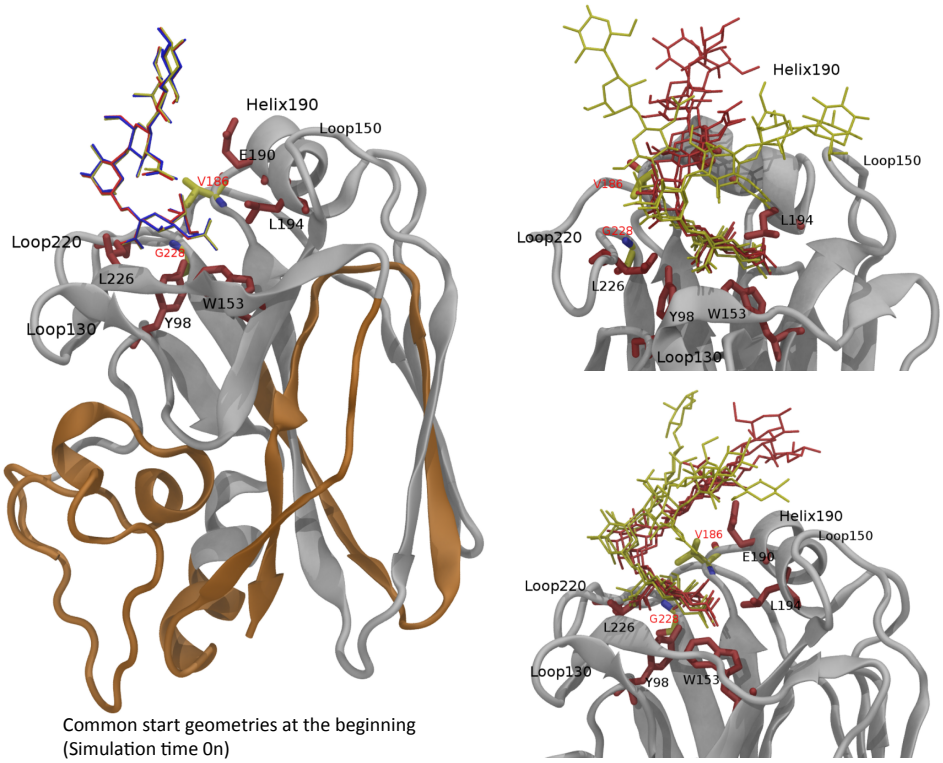


Fig. 13 LSTc-H7N9 complexes (wt and dm) at 41, 51 and 61 ns are superposed by the CA backbone in two different orientations (right). The three different ligand snapshots are shown in yellow (LSTc-H7N9wt) and in red (LSTc-H7N9dm). The orange ribbon shows the protein backbone restrained by harmonic potential during the MD simulation, while the white ribbon indicate protein backbone not restrained, i.e. free to adjust around the glycan

In light of these considerations, MD simulation predicts for LSTc a significant modification in the binding map when binding to H7 upon the considered mutations and also a reduced mobility, corresponding to a stronger binding.

Fig. 14 reports the distribution (histograms) of the distances between the methyl hydrogens of GlcNAc and the hydrogens belonging to the H7 RBS, for the three simulated complexes: LSTc-H7N9wt (red), LSTc-H7N9sm (green) and LSTc-H7N9dm (blue) calculated to the second half of each simulation trajectory.

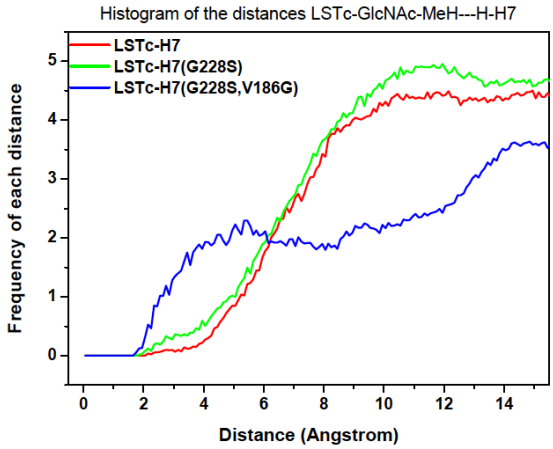


Fig. 14 The h-h histogram is calculated between 50ns until the end of the MD simulation trajectory, using a step = 2 geometries and a selection (protein and hydrogen). The distances between GlcNAc of LSTc and hydrogen of H7N9 are show in red for the wt, in green for the sm and in blu for the dm.

The histogram of the distances (Fig. 14) shows a population increase between 2 and 6 Å in going from H7N9wt to H7N9sm and more significantly to H7N9dm, corresponding to a prediction of an increase of the binding interaction after the first but more significantly the second mutation. These results are partially confirmed by the STD experiments reported in fig 7-right, where is possible to see the GlcNAc methyl hydrogens resonance of LSTc interacting with H7 that become visible in the single or in the double mutants. The effect of the mutations seems to slightly decrease the mobility of LSTc in going from H7N9wt to H7N9sm and to H7N9dm.

The same mutation effects on the recognition ability of H7 are also considered. Also in the case of LSTa-H7N9sm complex, the mutation of Gly228Ser introduces a hydrogen bond at the bottom of the RBS (Fig. 15), reinforcing the interaction between Neu5Ac and the bottom of RBS of HA.

In the double mutant of H7, where in addition to the mutation at position 228, a glycine is replaced by valine at position 186, reducing the steric hindrance between helix 190 and loop220 in the RBS of HA, the binding epitope of LSTa changes when it interacts with H7N9dm in comparison with H7N9sm and H7N9wt.

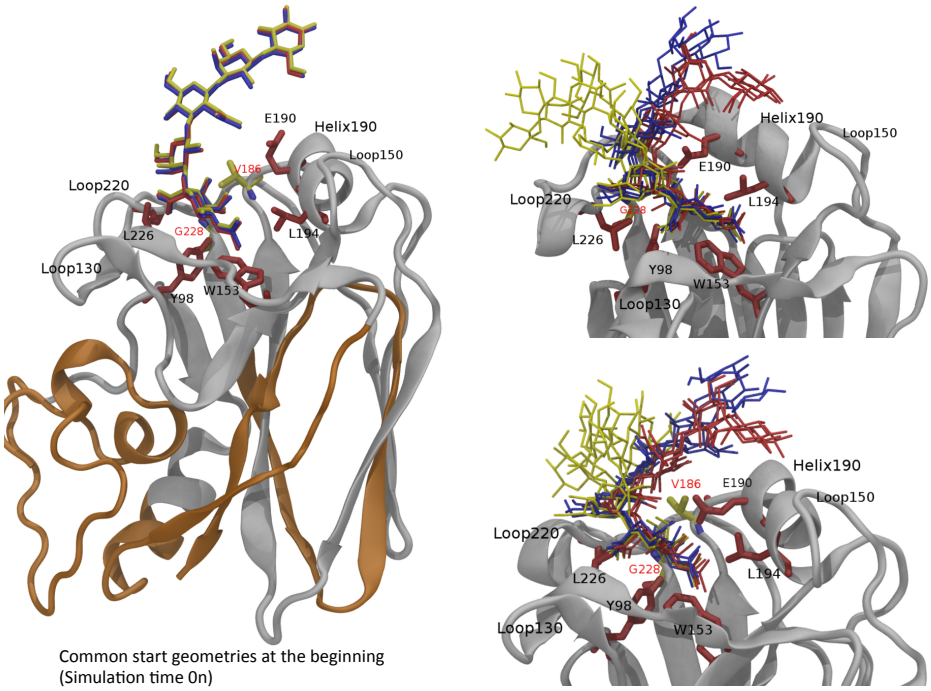


Fig. 15 LSTa-H7N9 complexes (wt and dm) at 36, 46 and 56 ns are superposed by the CA backbone in two different orientations (right). The three different ligand snapshots are shown in yellow (LSTa-H7N9wt), in blu (LSTa-H7N9sm) and in red (LSTc-H7N9dm). The orange ribbon shows the protein backbone restrained by harmonic potential during the MD simulation, while the white ribbon indicate protein backbone not restrained, i.e. free to adjust around the glycan

The mutation effects on LSTa-H7N9 binding interaction can be visualized drawing the distances between GlcNac and the Glu190 (Fig. 16), in which the

distance between GlcNAc and H7N9dm is smaller compared to the same distance observed in the complexes with H7N9wt and H7N9sm (Fig. 16).

The reducing end of LSTa changes significantly its contacts as a consequence of the mutations in the RBS of HA, whereas the non-reducing end remains almost unvaried. The features of LSTa binding to H7N9 and its mutant forms are in agreement with the corresponding STD NMR experiments, where data from all complexes indicate that Neu5Ac is the major residue involved in the interaction and, as observed previously for LSTc, its binding map in term of hydrogen resonances does not change upon mutation (Fig. 17).

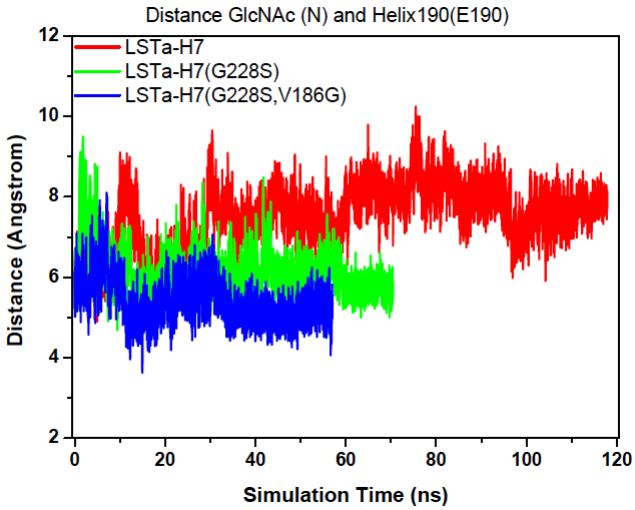


Fig. 16 Distances between GlcNAc and helix190 of H7N9wt in red, in green H7N9sm and in blu H7N9dm. The distances are defined between the GlcNAc nitrogen atom and the carboxyl atom of Glu190 side chain. The MD simulations of the sm and dm are considered preliminary data because the timescale explored is significantly smaller compared to the reference system LSTa-H7N9wt.

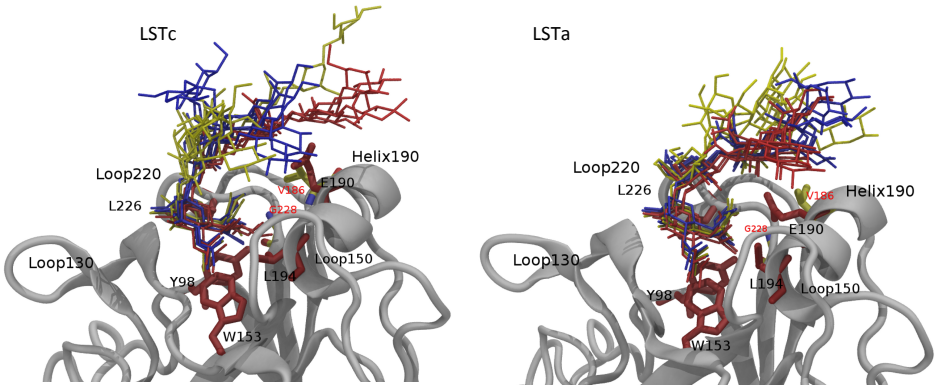


Fig 17 (Left) three snapshots of the LSTc-H7 complexes at 41, 51 and 61 ns are superposed by the CA backbone showing in particular the position of the Neu5Ac. The three different ligand conformations are shown in yellow (LSTc-H7wt) blue (LSTc-H7N9sm) and red (LSTc-H7N9dm). (Right) three snapshots of the LSTa-H7 complexes at 36, 46 and 56 ns are superposed by the CA backbone showing in particular the position of the Neu5Ac. The three different ligand conformation are shown in yellow (LSTa-H7wt), blue (LSTa-H7N9sm) and red (LSTa:H7N9dm)

The reducing end of LSTa when interacting with H7N9dm is closer to the RBS compare to the wt complex, in agreement with STD experiments which showed a weak signal corresponding to the resonance of H2 of Glc in the spectrum of LSTa-H7N9dm complex (Fig. 5).

Comparison between LSTc with H1N1(SC18) and H7N9wt. The comparison of the interaction between the pandemic virus H1N1(SC18) and H7N9wt with the human receptor LSTc, could help to define differences of recognition and interaction between two viruses which could be correlated in differences in term of virulence and transmissibility.

When LSTc binds HA, the interactions with SC18 and H7N9wt are very different (Fig. 18). The interaction between LSTc and H7N9wt occurs principally through Neu5Ac, on the contrary when LSTc binds SC18, the STD NMR spectrum (Fig.

18top) shows signals belonging to Gal1 (H1, H4, H5 and H6), GlcNAc(CH_3), Gal2 (H4) and Glc (H5 and H6).

In H7N9wt, loop150 includes two extra amino acids that make it longer compared to SC18 and causes the protrusion of the 150-loop into one edge of the binding site where it has the potential to influence receptor specificity (Fig. 18 bottom). In fact the variation of the loop150 changes significantly the binding epitope of LSTc when interacts with SC18 compared to H7N9 (in agreement with Xiong et al. [3]). As a matter of fact the presence of this longer 150-loop in H7N9 does not allow LSTc to interact optimally with the helix190 as in SC18.

The interaction between the two complexes is quite different also in the distance between the methyl group of GlcNAc and the RBS of HA, which is shorter in LSTc-SC18 compare to LST-H7N9wt complex.

This result is clearly supported by the STD NMR data (Fig. 18 top), in which the signal of the CH_3 of the GlcNAc is absent in the STD spectrum of LSTc-H7N9wt, whereas is quite strong in the STD of LSTc-SC18 complex.

The presence of a leucine in position 226 in H7N9, instead a glutamine as in SC18, introduces a hydrophobic space between loop220 and loop130, increasing their distance in H7N9 compared to SC18. The difference in the LSTc binding maps during the interaction with H7N9 and SC18 correlates with different measurable binding affinity, that for SC18 is significantly stronger than H7N9, as measured by Shi et al. [4] (Fig.1 A and Fig. 1C).

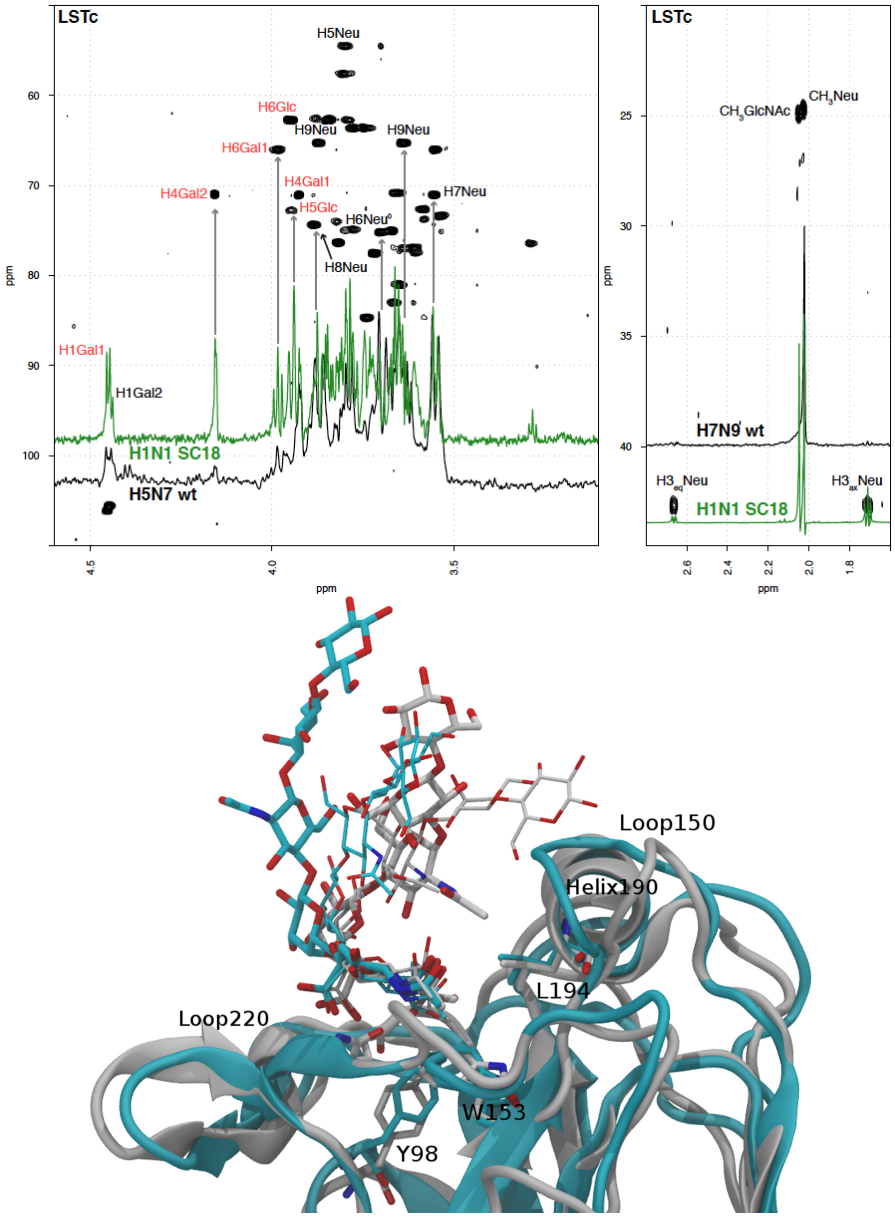


Fig. 18 On the top in black the STD spectrum of LSTc-H7N9wt complex while in green the STD spectrum of LSTc-SC18 complex superimpose to the HSQC spectrum (in black) of LSTc. On the bottom superimposition by protein backbone (type CA, and resid 2 to 200) of LSTc-SC18 (grey) and LSTc-H7N9wt (cyan), simulation time = 100ns. Residual RMSD = 6.144 Å. Thinner lines draw LSTc at intermediate time (simulation time = 50ns) binding to H7 (cyan) and SC18 (grey).

CONCLUSION

Combining NMR experiments and MD simulation analysis we have obtained an overview of the interaction between LSTc and LSTa with H7N9wt and its two mutant forms, with the idea to find if these mutations may switch H7 binding preference from the original avian (LSTa type) to the human (LSTc type) cell surface receptors.

In this case, the single (Gly228-Ser) or double mutant (Gly228-Ser, Val186-Gly) does not change dramatically the interaction with both human and avian receptor models. In both glycans receptors the interaction occurs principally through the Neu5Ac, as shown in the model proposed by the MD simulations and confirmed by STD NMR experiments. The double mutant of H7 seems to slightly increase the binding with both LSTc and LSTa, because in both cases the reducing end of the glycan receptors is closer to the protein in comparison to the wild type taken as binding epitope reference. Notably, the major difference in term of interaction between human or avian glycan receptors is due to the reducing end, in which in LSTc shows grater mobility than LSTa.

All these data indicate that these two mutations in H7N9 are not crucial to switch preferences from avian to human receptors.

This study also permit to compare the differences of interaction between two different hemagglutinin proteins, specifically the pandemic H1N1 (SC18) virus and the less pandemic H7N9wt virus, with the human glycosidic receptor model (LSTc), in order to understand how different classes of viruses recognize the same glycan receptor.

This study requires further investigations, first of all to complete the MD simulations of all the complexes and to perform analysis upon them, and to

perform some competition and titration experiments using different concentrations of ligands using further amounts of glycan receptors. Another interesting study could involve some *in silico* mutations in H7N9, with the aim to find new indication for point-like mutations able switch the specificity of H7 from the original avian (α2-3) to the human (α2-6) glycan receptor, corresponding to a requisite for an avian virus to become pandemic.

EXPERIMENTAL PROCEDURES

Cloning, Baculovirus Synthesis, and Mammalian Expression and Purification of HA. Anh13 WT, Gly228-Ser and Gly228-Ser/Val186-Gly sequences were codon optimized for mammalian expression, synthesized (DNA2.0, Menlo Park, CA), and subcloned into modified pcDNA3.3 vector for expression under CMV promoter. Recombinant expression of HA was carried out in HEK293-F FreeStyle suspension cells (Invitrogen, Carlsbad, CA) cultured in 293-F FreeStyle Expression Medium (Invitrogen, Carlsbad, CA) manteined at 37° C, 80% umidity, and 8% CO_2. Cells were transfected with poly-ethylene-imine Max (PEI-MAX, PolySciences, Warrington, PA) with the HA plasmid and were harvested 7 days post-infection. The supernatant was collected by centrifution, filtered through a 0.45 μm filter system (Nalgene, Rochester, NY), and supplemented with 1:1,000 diluted protease inhibitor cocktail (Calbiochem filtration and supplemented with 1:1,000 diluted protease inhibitor cocktail (EMD Millipore, Billerica, MA)). HA was purified from the supernatant using His-trap columns (GE Healtcare) on an AKTA Purifier FPLC system. Eluting fractions containing HA were pooled, concentrated, and buffer exchanged into 1 x PBS (pH 7.4) using

100 kDa MWCO spin columns (Millipore). The purified protein was quantified using the BCA method (Pierce, Rockford, IL).

NMR Analysis of H7N9wt, sm and dm with LSTc and LSTa. Saturation Transfer Difference (STD) samples were prepared washing the proteins H7N9wt, sm and dm (1 mg/mL) with buffered solution (150 mM sodium chloride, 100 mM sodium phosphate, 0.3 mM d- EDTA, D2O, pH7.2 (Sigma Aldrich)) using Amicon® Ultra centrifugal filters, 10 KDa membrane (Millipore). Each ligand (LSTc or LSTa) were added to the corresponding protein sample reaching a final molar ratio of 100:1 [glycan receptor:HA] for the STD measurements. For the competitive experiment to a 200 μg of the first ligand (100:1), another 200 μg of the second ligand was added to the NMR tube. The protein concentration for the STD measurements was 0.01 mM. NMR spectra were acquired using a Bruker 600 and 900 MHz AVANCE series NMR spectrometer, both equipped with a high sensitivity 5 mm TCI cryoprobe. LSTc and LSTa resonances could be found in fig. 2 and table 1 chapter 2 . For the STD experiments the on-resonance frequency was set at 7.3 ppm [6600 Hz] and the off-resonance frequency at 20.0 ppm [18000 Hz], a train of 40 gaussian-shaped pulses of 50 ms each were applied to produce a selective saturation of 2 s and D1 was set to 6 s. The number of scans was 1K or 2K and the spectral width used was 12626 Hz. The spectra were recorded at 295 K.

Molecular Dynamics Simulations The structures 4BSE and 4BSF of the Protein Data Bank include a single monomer of the H7 (AH-H7N9) protein made by chain A and B, with a co-crystallized trisaccharides resolved structure of LSTc

and LSTa in H7 principal RBS. The trisaccharides resolved structure in 4BSE and 4BSF (Neu5Ac-Gal-GlcNAc) correspond to the three residues starting from the non-reducing end of LSTc and LSTa respectively. The trisaccharide-H7 complex structures in 4BSE and 4BSF are used as *template* to build the LSTc:H7N9wt and LSTa:H7N9wt model complexes that are at the base of the investigation on H7N9 recognition ability toward LSTc (human) and/or LSTa (avian) glycan receptors. The LSTc and LSTa penthasaccharide conformations taken as the most probable in solution (Sassaki et al.[7]), and they are superposed by their Neu5Ac residues on the corresponding co-crystallized trisaccharides in 4BSE and 4BSF structure respectively; the RSMD distance after superposition of the Neu5Ac residue are 3.32 and 3.34Å respectively. The LSTc-H7N9wt and LSTa-H7N9wt models structures were built merging the 3D structure of LSTc or LSTa glycan with the RBS structure of H7, corresponding to the chain A sequence between L60 and L260 (H3 numbering) extracted from 4BSE and 4BSF respectively. The software used for this molecular editing was VMD. The H7 mutated complexes including the single mutant H7N9sm (H7G228S) and the double mutant H7N9dm (Gly228Ser, Val186Gly) were obtained applying the *"in silico"* mutation (MAESTRO9.8 graphical interface) on the RBS of the two previously built complexes, generating four additional complexes: LSTc-H7N9sm and LSTc-H7N9dm, with the corresponding including the avian glycan receptor: LSTa-H7N9sm and LSTa-H7N9dm. Ambertools 1.4 was used to build the force field. GLYCAM06/Amber was used to describe the glycan and protein part of the complexes. The simulation cell was built by enveloping each macromolecule by a water layer (TIP3P water model) 15 Å wide in three directions, resulting in an orthogonal cell with an edge of approximately 100 Å. The non bonded potential energy was described using the standard cutoff (12

Å) technique for both electrostatic and dispersive interactions. Each cell was minimized using 100 K steps of the default minimization algorithm included in NAMD. Then 1 ns of MD simulation sampling the NPT ensemble was used to equilibrate the cell density. The simulation temperature was set at 295 K and maintained by a Langevin thermostat as implemented in NAMD, while the Nosé-Hoover Langevin piston algorithm controlled the pressure (1.01325 bar) applied to the cell walls. During the minimization and cell density equilibration steps, a harmonic potential energy restraint (harmonic constant of 50 kcal mol−1) on all the atoms of the complex was applied, while the water molecules were allowed to move freely. The MD simulation for all the modeled complexes was approximately 100 ns, and was completed leaving the H7 sequence surrounding the glycan (35-50 and 70-173 A.A. residues in H7 RBS) free to move. Soft harmonic restraint on the HA backbone atoms (Cα, N, and carbonyl carbon) with harmonic constant of 2.0 kcal mol^{-1}, were applied on the remain sequences, to maintain as much as possible the II and III structure of the H7 RBS.

REFERENCES

[1] R. Gao, B. Cao, Y. Hu, Z. Feng, D. Wang, W. Hu, J. Chen, Z. Jie, H. Qiu, K. Xu, X. Xu, H. Lu, W. Zhu, Z. Gao, N. Xiang, Y. Shen, Z. He, Y. Gu, Z. Zhang, Y. Yang, X. Zhao, L. Zhou, X. Li, S. Zou, Y. Zhang, X. Li, L. Yang, J. Guo, J. Dong, Q. Li, L. Dong, Y. Zhu, T. Bai, S. Wang, P. Hao, W. Yang, Y. Zhang, J. Han, H. Yu, D. Li, G. F. Gao, G. Wu, Y. Wang, Z. Yuan, and Y. Shu, "Human infection with a novel avian-origin influenza A (H7N9) virus.," *N. Engl. J. Med.*, vol. 368, no. 20, pp. 1888–97, May 2013.

[2] Q. Zhang, J. Shi, G. Deng, J. Guo, X. Zeng, X. He, H. Kong, C. Gu, X. Li, J. Liu, G. Wang, Y. Chen, L. Liu, L. Liang, Y. Li, J. Fan, J. Wang, W. Li, L. Guan, Q. Li, H. Yang, P. Chen, L. Jiang, Y. Guan, X. Xin, Y. Jiang, G. Tian, X. Wang, C. Qiao, C. Li, Z. Bu, and H. Chen, "H7N9

influenza viruses are transmissible in ferrets by respiratory droplet.," *Science*, vol. 341, no. 6144, pp. 410–4, Jul. 2013.

[3] X. Xiong, S. R. Martin, L. F. Haire, S. a Wharton, R. S. Daniels, M. S. Bennett, J. W. McCauley, P. J. Collins, P. a Walker, J. J. Skehel, and S. J. Gamblin, "Receptor binding by an H7N9 influenza virus from humans.," *Nature*, vol. 499, no. 7459, pp. 496–9, Jul. 2013.

[4] Y. Shi, W. Zhang, F. Wang, J. Qi, Y. Wu, H. Song, F. Gao, Y. Bi, Y. Zhang, Z. Fan, C. Qin, H. Sun, J. Liu, J. Haywood, W. Liu, W. Gong, D. Wang, Y. Shu, Y. Wang, J. Yan, and G. F. Gao, "Structures and receptor binding of hemagglutinins from human-infecting H7N9 influenza viruses.," *Science*, vol. 342, no. 6155, pp. 243–7, Oct. 2013.

[5] K. Tharakaraman, A. Jayaraman, R. Raman, K. Viswanathan, N. W. Stebbins, D. Johnson, Z. Shriver, V. Sasisekharan, and R. Sasisekharan, "Glycan receptor binding of the influenza A virus H7N9 hemagglutinin.," *Cell*, vol. 153, no. 7, pp. 1486–93, Jun. 2013.

[6] C. McCullough, M. Wang, L. Rong, and M. Caffrey, "Characterization of influenza hemagglutinin interactions with receptor by NMR.," *PLoS One*, vol. 7, no. 7, p. e33958, Jan. 2012.

[7] G. L. Sassaki, S. Elli, T. R. Rudd, E. Macchi, E. A. Yates, A. Naggi, Z. Shriver, R. Raman, R. Sasisekharan, G. Torri, and M. Guerrini, "Human (alfa2->6) and avian (alfa2->3) sialylated receptors of influenza A virus show distinct conformations and dynamics in solution," *Biochemistry*, vol. 52, no. 41, pp. 7217–7230, 2013.

Conclusions and Remarks

GENERAL CONCLUSIONS

This work provides new insights into the detailed description of the conformation of the human (LSTc) and avian (LSTa) glycan receptors of the influenza A viruses free in solution and upon binding with different hemagglutinins.

The aim of the first part of the project was the characterization of the dynamic and conformation of the two glycans in solution. Detail study involving NMR, MD and numerical analysis of the data generated provided information regarding the different conformations acquired from the LSTc and LSTa and also the dynamic properties of both the receptors. While LSTc, in solution, assumes preferentially a bent shape characterized by a topological θ angle (which is the angle between the three anomeric carbons of Neu5Ac,GlcNAc and Gal) of 87°; LSTa presents a more linear conformation with a θ angle of 159°. These structural peculiarities result in great differences in term of dynamic: LSTc has the Neu5Ac that moves in tandem with the N-acetyl group of GlcNAc, and its movement is independent of the reducing end, whereas LSTa has a movement similar to a flexible bow that has the reducing and non-reducing disaccharides that move around the GlcNAc fulcrum. This is also due to the intricate hydrogen bond network between GlcNAc and the non-reducing end (Neu5Ac) of LSTc; these hydrogen bonds are not present in LSTa, which exhibits only interactions between directly adjacent residues.

The interaction between LSTc and LSTa respectively and three different hemagglutinins, SC18 derived from the virus H1N1 (A/South Carolina/1/18), NY18 derived from SC18 with a single Asp225-Gly mutation; and AV18, a

double mutant of SC18 (Asp225-Gly/Asp190-Glu), were investigated. On the basis of the NMR STD experiments and MD simulations, it was demonstrated that mutations of HA RBS imposed different constrains on the conformation and topology of the bound glycan receptors, governing the specificity and affinity of the binding. More specifically, the receptor binding site (RCB) of SC18 and NY18 imposes constrains on human receptor (LSTc) such that the glycan assumes exclusively an umbrella-like topology, characterized by a θ angle <100°, whereas RCB of AV18 predominantly samples a cone like topology for both human and avian receptor, characterized by θ angle >100°. Moreover, the bound human receptor exhibits restricted glycosidic torsional angles compared to the free state when bound to SC18 and NY18, which are distinct from those observed on the double mutant AV18. Even to lesser extent, these restrictions were observed also in the LSTa-AV18 complex, whereas in LSTa-NY18 complex all glycosidic torsional angles are similar to those seen in the free state. Results indicated that even one or two amino-acid changes in the HA RBS impose different constrains on the conformation of the bound glycan which are strictly related to the specificity and affinity of the binding (LSTc SC18>NY18>>AV18; LSTa AV18>NY18>>SC18).

During the last part of the study, the investigation was focused on the characterization of the interaction between human and avian glycan receptors and the new H7N9 influenza A virus strain, which emerged in 2013, and two of its mutants. The binding epitope of both ligands for each complex tested were described. The interaction in each complex occurs principally through the Neu5Ac, and in these cases the mutations do not change dramatically the interaction. STD-NMR and MD results obtained for the LSTc-H7N9 complex

compared to the more pandemic H1N1 (SC18) virus showed a different pattern of interaction, the former interacting mainly with the Neu5Ac residue. H7N9 has a 130 loop longer compared to the same loop of SC18. This longer loop, pushing away the reducing end of LSTc, shrinks the conformational constrains of the glycan imposed by SC18, correlated with the lower affinity of the virus. Preliminary NMR and MD experiments showed that Gly228Ser and Gly228Ser/Val186Gly mutations in H7N9 do not affect the specificity of the virus, since they did not show any preference for human and avian receptor.

The future developments of this project will regard the *in silico* generation of mutant forms of H7N9 in order to identify mutation/s able to switch completely the preference from avian to human receptor and in this way characterize the conformation and topology of interaction between the RBS of hemagglutinins which present high affinity and specificity toward human receptor.

RIASSUNTO

Le interazioni proteine-glicani regolano numerosi processi biologici, uno di questi è il processo d'infezione ad opera del virus dell'influenza A. Il virus dell'influenza A è un virus a RNA formato da otto geni, tre dei quali esprimono proteine - emoagglutinina (HA), neuraminidasi (NA) e polimerasi (PB) –, che risultano particolarmente critiche nell'infezione e nella trasmissione uomo-uomo[1]. L'infezione inizia con il legame dell'HA, una proteina espressa sulla superficie del virus, ai recettori glicanici presenti sulle cellule dell'ospite; questa interazione è governata dal tipo di legame (α2→6 o α2→3) tra l'acido sialico (Neu5Ac) e il galattosio (Gal) dell'estremità terminale non riducente del recettore.

I recettori glicanici umani, siti di riconoscimento per i virus human-adapted, sono espressi principalmente nel tratto superiore dell'epitelio respiratorio e presentano un legame α2→6 tra l'acido sialico e il galattosio nell'estremità non riducente. I virus aviari, presenti nel tratto inferiore dell'epitelio respiratorio umano o nell'intestino degli uccelli, riconoscono invece recettori glicanici che esprimono un legame α2→3 tra l'acido sialico e il galattosio. Per cambiare specificità (da aviario a umano) un virus deve quindi modificare la sua preferenza da recettori glicanici legati α2→3 (aviari) a quelli legati α2→6 (umani)[2].

Tramite analisi di strutture a raggi X e studi di dinamica molecolare su complessi glicano-HA è stato dimostrato che la conformazione e la topologia dei recettori glicanici, ricoprono un ruolo chiave nella regolazione della specificità e dell'affinità di queste interazioni[2],[3].

In aggiunta alle informazioni ottenute dalla cristallografia, l'applicazione di tecniche NMR, quali l'STD (Saturation Transfer Difference), possono fornire informazioni relative alla dinamica del recettore in soluzione e alle eventuali transizioni conformazionali del recettore causate dall'interazione con l'HA del virus.

Nella prima fase del lavoro ci siamo occupati di caratterizzare in dettaglio la dinamica e la conformazione in soluzione di due pentasaccaridi, usati come modelli per il recettore aviario (LSTa, Neu5Ac-α(2→3)-Gal-β(1→3)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc) e umano (LSTc, Neu5Ac-α(2→6)-Gal-β(1→4)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc), utilizzando tecniche NMR (Nuclear Magnetic Resonance) e simulazioni di dinamica molecolare (MD).

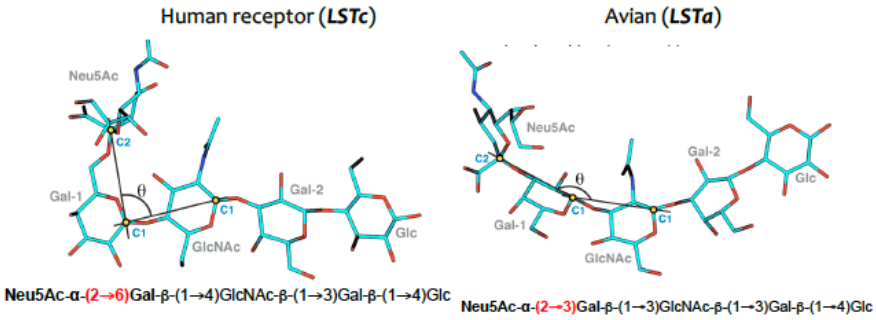


Fig. 1 Struttura rispettivamente di LSTc ed LSTa. L'angolo θ è definite dagli atomi C2 di Neu5Ac, C1 di Gal1 e C1 di Glc NAc rispettivamente, andando dall'estremità non-riducente a quella riducente.

Abbiamo iniziato caratterizzando entrambi i recettori (LSTc ed LSTa) in soluzione, utilizzando esperimenti NMR mono e bi-dimensionali, quali: ^{1}H, COSY, TOCSY, HSQC, HSQC-TOCSY ed HMBC; ottenendo in questo modo l'assegnamento completo dei due pentasaccaridi. A questo punto la caratterizzazione conformazionale dei due pentasaccaridi in soluzione è stata

fatta tramite esperimenti selettivi e di 2D-NOESY, grazie ai quali è stato possibile stimare anche le distanze protone-protone e confrontare quindi i dati sperimentali con quelli teorici, ottenuti tramite dinamica molecolare. La presenza di un segnale NOE tra l'H5 di Neu5Ac e il CH_3 di GlcNAc nello spettro NOESY di LSTc ha consentito di confermare la conformazione ripiegata di LSTc.

L'assenza di questo segnale nello spettro NOESY di LSTa, suggerisce invece una conformazione lineare per quest'ultimo oligosaccaride. Queste peculiarità conformazionali, caratteristiche dei due recettori, potrebbero essere correlate alla specificità dell'interazione tra il recettore glicanico e l'emoagglutinina.

Un parametro importante per descrivere le conformazioni dei glicani è l'angolo θ, definito dall'angolo compreso tra i tre carboni anomerici successivi, partendo dal residuo di Neu5Ac dell'estremità non-riducente dell'oligosaccaride (Fig.1). I NOE sperimentali, hanno confermato il modello ottenuto tramite la dinamica molecolare[3], che ha permesso di dimostrare che mentre per LSTa θ assume preferenzialmente valori intorno a 160°, LSTc presenta un valore di θ intorno a 90°, confermando quando precedentemente osservato [4].

La seconda parte del lavoro è stata focalizzata allo studio dell'interazione tra HAs (wild type e mutate) del virus dell'influenza A con i recettori glicosidici umano e aviario; con lo scopo di capire come pochissime mutazioni all'interno dell'emoagglutinina del virus, siano in grado di cambiare la specificità del virus, ed anche la sua trasmissibilità.

Nel 1918 "la spagnola", come fu soprannominata all'epoca la pandemia causata dal virus dell'influenza aviaria H1N1, fu di profonda importanza storica e geopolitica, poiché circa 50 milioni di persone morirono a seguito

dell'infezione, che fece più morti della stessa prima guerra mondiale: una sua caratteristica chiave fu l'elevata trasmissibilità.

Nel 2009, l'emergenza data dall'arrivo di un nuovo virus H1N1 pose fortemente la questione circa il rischio di una nuova pandemia a seguito dell'evoluzione di un virus in una variante epidemica; alcuni di questi sotto-tipi sono considerati avian-adapted poiché derivano da virus che infettano gli uccelli ma sono in grado di trasferirsi anche all'uomo.

La relazione tra la specificità dell'HA verso i recettori e la trasmissibilità del virus è stata precedentemente dimostrata usando il prototipo del virus SC18 (H1N1) A/South Carolina/1/1918. Mentre SC18 mostra un'efficienza di trasmissione molto alta nei furetti (usati come modello per studiare la trasmissibilità aerea del virus), una singola mutazione (Asp225-Gly) nell'HA risulta in un virus (NY18) che presenta invece un'efficienza di trasmissione bassa. Infine una seconda mutazione nell'HA (Asp190-Glu e Asp225-Gly) produce un virus (AV18) che non è più in grado di trasmettersi. Studi precedenti hanno inoltre dimostrato che mentre le emoagglutinine di SC18 e AV18 mostrano un'alta affinità rispettivamente verso il recettore umano e aviario, NY18 lega entrambi i recettori, ma con bassa affinità[1][3][4].

Utilizzando esperimenti NMR di STD, si può ottenere l'epitope mapping della porzione del recettore che interagisce con l'emoagglutinina. Questa tecnica è stata quindi utilizzata per studiare l'interazione tra SC18, NY18 ed AV18 con LSTa ed LSTc. Oltre a confermare la specificità di interazione del recettore umano e aviario, già definite da test biochimici, mediante esperimenti STD, è stato inoltre possibile definire, anche per il recettore aviario, l'epitope mapping dell'interazione, dimostrando come LSTa interagisca con NY18 ed AV18 quasi

esclusivamente tramite il Neu5Ac[5]. Al contrario, LSTc interagisce con SC18 e NY18 sia tramite Neu5Ac che con Gal1 e GlcNAc.

Per verificare se l'interazione tra LSTc e SC18 o NY18 comporta variazioni conformazionali, sono stati utilizzati esperimenti di tr-NOESY. Il segnale NOE tra l'H5 di Neu e il CH_3 di GlcNAc, osservato nello spettro dell'oligosaccaride in soluzione, è presente anche nello spettro tr-NOESY del complesso LSTc/NY18, indicando che la conformazione ripiegata del ligando viene mantenuta all'interno del complesso. Nello spettro tr-NOESY del complesso LSTc/SC18, è possibile osservare un segnale NOE aggiuntivo tra il CH_3 di GlcNAc e l'H8/H9 di Neu5Ac, compatibile con una variazione conformazionale di LSTc a seguito dell'interazione con SC18, dovuta ad una riduzione della distanza tra GlcNAc e Neu5Ac.

L'identificazione dei contatti che permettono l'interazione del mutante NY18 con entrambi i recettori rappresenta un punto chiave per studiare il meccanismo tramite il quale il virus dell'influenza A è in grado di adattarsi e passare dal legame con il recettore aviario a quello umano. Partendo dalle strutture ottenute tramite raggi X, abbiamo dunque utilizzato la dinamica molecolare allo scopo di descrivere l'interazione tra recettori glicanici ed emoagglutinina, e vedere come queste interazioni cambiano a seguito delle mutazioni. Il complesso LSTc/SC18 è stato costruito a partire dalla struttura co-cristallizzata di SC18 con LSTc (2WRG), a cui abbiamo sovrapposto la struttura di LSTc in soluzione ottenuta precedentemente[6]. I complessi con i mutanti, sono stati invece ottenuti partendo dal complesso LSTc-SC18 inserendo la singola e doppia mutazione.

In maniera analoga sono stati costruiti i complessi LSTa-AV18 ed LSTa-NY18. In questo modo tutti i complessi presentavano la stessa geometria di partenza

all'inizio della simulazione di dinamica molecolare, garantendo così la confrontabilità tra le varie strutture alla fine della stessa. Durante le simulazioni tra LSTc e HA, si nota come la distanza media tra LSTc e i residui più vicini del sito attivo dell'emaggluininina sia minore nel complesso LSTc-SC18 rispetto al complesso LSTc-NY18 (Fig. 2a e 2b).

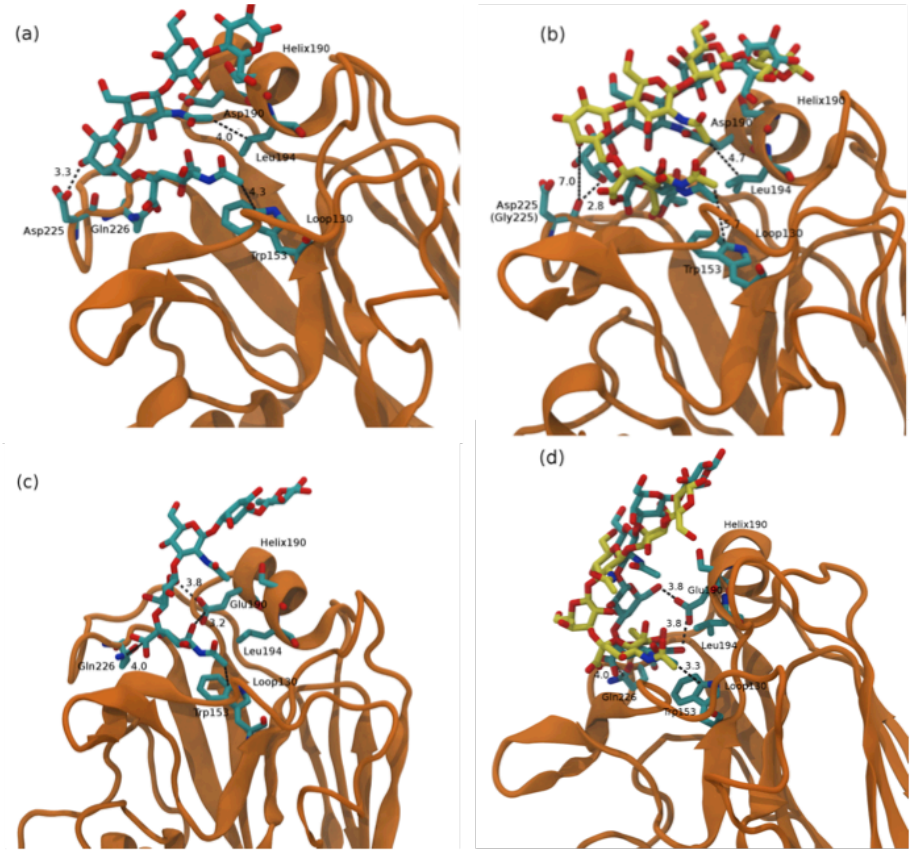


Fig. 2 (a) Struttura del complesso LSTc-SC18 (b) Sovrapposizione dei RBSs' di LSTc-SC18 (cyan) con LSTc-NY18 (giallo) (c) Struttura del complesso LSTa-AV18 (d) Sovrapposizione delle strutture dei complessi LSTa-AV18 (cyan) e LSTa-NY18 (giallo).

La simulazione ottenuta per LSTc-NY18 mostra infatti che la mutazione in posizione 225 risulta in una riduzione delle interazioni elettrostatiche sul fondo del sito di legame del recettore (Fig. 2b), mentre la doppia mutazione in AV18 introduce una catena laterale più ingombrante nell'elica 190, riducendo in questo modo l'interazione tra LSTc e l'elica stessa.

Nelle simulazioni di dinamica molecolare tra i complessi LSTa-HA, è invece possibile osservare che la distanza media tra LSTa e AV18 è minore in confronto al complesso LSTa-NY18 (Fig. 2d). Questo è dovuto al fatto che la catena laterale del Glu190 in AV18 è posizionata in maniera ottimale per interagire con LSTa, in confronto alla catena più corta dell'Asp presente in NY18.

Sembra dunque che Il virus pandemico SC18, a confronto con il singolo (NY18) e doppio (AV18) mutante, imponga vincoli conformazionali più stringenti al recettore umano, proprietà che sembra essere correlata all'affinità dell'interazione recettore-HA, misurata anche tramite saggi biochimici[5].

L'angolo θ per LSTc legato ad HA cresce infatti andando da SC18 verso AV18 (Fig. 3). Questa relazione tra affinità e vincoli conformazionali imposti al recettore, è stata osservata anche per il virus aviario-adattato AV18, il quale impone vincoli più stringenti al recettore aviario in confronto a quelli imposti da NY18.

Durante il 2013, una nuova forma virale d'influenza A si è diffusa in Cina. Il nuovo virus, H7N9, deriva da un riassortimento: un processo in cui due o più virus influenzali co-infettano un singolo ospite e scambiano così geni. Isolando due virus, A/Anhui/1/2013 e A/Shanghai/1/2013 è stato osservato [7][8] come il primo possedesse la mutazione Gln226Leu, associata solitamente ad un'affinità maggiore verso il recettore umano.

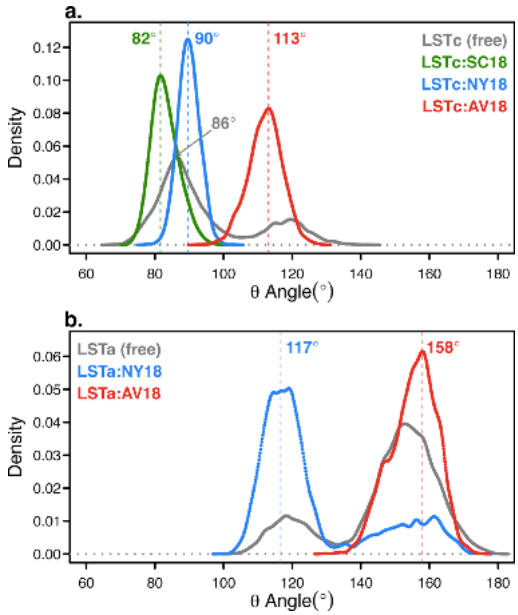


Fig. 3 (a) Valori di angolo θ per LSTc in soluzione (grigio), LSTc-SC18 (verde), LSTc-NY18 (blu) ed LSTc-AV18 (rosso) (b) Valori di angolo θ per LSTa in soluzione (grigio), LSTa-NY18 (blu) e LSTa-AV18 (rosso).

Nell'ultima parte del nostro lavoro tramite l'impiego di approcci complementari, quali NMR e dinamica molecolare, abbiamo così esteso il nostro studio a un virus meno pandemico, H7N9, e a due suoi mutanti, allo scopo di caratterizzarne l'interazione con LSTc ed LSTa. Precedenti studi hanno evidenziato come questo nuovo virus, A/Anhui/1/2013 (H7N9wt) è in grado di legare sia il recettore umano che aviario, mentre una singola mutazione Gly228-Ser in HA sembra in grado di aumentare l'interazione verso il recettore umano[9]–[11].

Utilizzando esperimenti STD abbiamo confermato che entrambi i recettori interagiscano con H7N9wt prevalentemente tramite il Neu5Ac. Partendo dalle strutture co-cristallizzate di LSTc con H7 di origine umana (4BSE) e di LSTa co-cristallizzato con H7 (4BSF)[12] sono stati generati i modelli di partenza per le

simulazioni di dinamica molecolare; da cui è stato possibile osservare come l'estremità non riducente di LSTc possieda una libertà conformazionale molto maggiore in confronto a LSTa. In maniera analoga sono stati studiati i complessi LSTc-H7N9 singolo mutante (Gly228-Ser), LSTc-H7N9 doppio mutante (Gly228-Ser, Val186-Gly), LSTa-H7N9sm e LSTa-H7N9dm. Il recettore umano è in grado di interagire con il singolo e doppio mutante, tramite Neu5Ac ed il CH_3 di GlcNAc, segnale che non è invece presente nello spettro STD del complesso LSTc-H7N9wt. I dati sperimentali sono in accordo con le simulazioni di dinamica molecolare, queste ultime mostrano delle fluttuazioni minori della distanza misurata tra il CH_3 di GlcNAc e l'elica 190 di H7N9sm in confronto con H7N9wt. Inoltre tramite dinamica si evidenzia come la sostituzione della serina con una glicina nella mutazione in posizione 228, introduca un ulteriore legame idrogeno sul fondo del sito di legame del recettore, rinforzando l'interazione tra l'H5 di Neu5Ac e H7N9sm. Esperimenti STD sono stati acquisiti anche con i complessi LSTa-H7N9sm e LSTa-H7N9dm. Nell'interazione con H7N9dm lo spettro STD mostra anche un segnale del glucosio, indicando la maggiore vicinanza dell'estremità riducente di LSTa all'elica 190 rispetto al wt.

Le strutture dei complessi LSTc-H7N9, ottenuti tramite dinamica mostrano come l'effetto delle mutazioni sembra ridurre leggermente la mobilità di LSTc, andando dall'interazione con H7N9wt, passando per H7N9sm ed infine diminuendo ulteriormente con H7N9dm. Sebbene i medesimi complessi interagiscano con LSTa essenzialmente con il Neu5Ac, le mutazioni rinforzano tali interazioni, coinvolgendo parzialmente anche alcuni residui dell'estremità riducente in H7N9dm.

L'ultima parte dello studio ha avuto come oggetto il confronto dell'interazione tra il recettore umano (LSTc) e due virus, H7N9wt e SC18, con caratteristiche

differenti a livello di trasmissibilità; allo scopo di capire come virus differenti e con diversa virulenza siano in grado di riconoscere e legare lo stesso recettore. Sulla base dei dati sperimentali e di dinamica molecolare, abbiamo dimostrato che LSTc interagisce con SC18 con una conformazione sostanzialmente diversa rispetto a quella che assume quando lega H7N9wt. Inoltre, l'interazione tra LSTc ed SC18 risulta più forte in confronto a quella con H7N9wt, dato confermato sia tramite esperimenti STD (numero maggiore di segnali STD nello spettro LSTc-SC18 rispetto a quelli presenti in LSTc-H7N9wt), sia dallo studio di dinamica molecolare (LSTc appare più vicino a SC18 rispetto a H7N9wt) (Fig. 4).

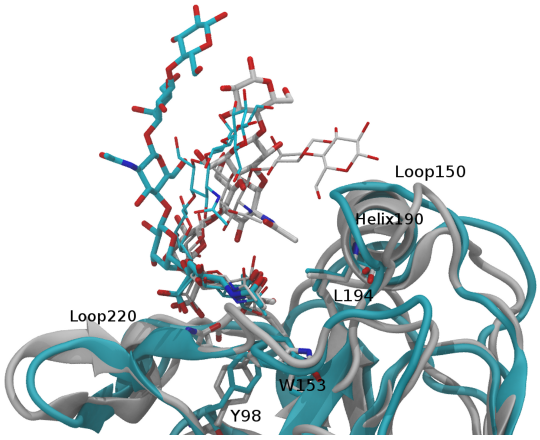


Fig. 4. Sovrapposizione tramite backbone dei complessi LSTc-SC18 (grey) e LSTc_H7N9wt (cyano)

In questo lavoro abbiamo descritto le basi strutturali dell'interazione tra l'emoagglutinina di nuovi virus e i recettori umani e aviari, combinando l'approccio sperimentale a tecniche computazionali. Pertanto questa metodologia potrà essere usata per predire la potenziale infettività di nuove varianti del virus dell'influenza e quindi come strumento utile per la sorveglianza di nuovi virus pandemici.

BIBLIOGRAFIA

[1] T. M. Tumpey, T. R. Maines, N. Van Hoeven, L. Glaser, A. Solórzano, C. Pappas, N. J. Cox, D. E. Swayne, P. Palese, J. M. Katz, and A. García-Sastre, "A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission.," *Science*, vol. 315, no. 5812, pp. 655–9, Feb. 2007.

[2] Z. Shriver, R. Raman, K. Viswanathan, and R. Sasisekharan, "Context-specific target definition in influenza a virus hemagglutinin-glycan receptor interactions.," *Chem. Biol.*, vol. 16, no. 8, pp. 803–14, Aug. 2009.

[3] A. Chandrasekaran, A. Srinivasan, R. Raman, K. Viswanathan, S. Raguram, T. M. Tumpey, V. Sasisekharan, and R. Sasisekharan, "Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin.," *Nat. Biotechnol.*, vol. 26, no. 1, pp. 107–13, Jan. 2008.

[4] D. Xu, E. I. Newhouse, R. E. Amaro, H. C. Pao, L. S. Cheng, P. R. L. Markwick, J. A. McCammon, W. W. Li, and P. W. Arzberger, "Distinct glycan topology for avian and human sialopentasaccharide receptor analogues upon binding different hemagglutinins: a molecular dynamics perspective.," *J. Mol. Biol.*, vol. 387, no. 2, pp. 465–91, Mar. 2009.

[5] S. Elli, E. Macchi, T. R. Rudd, R. Raman, G. Sassaki, K. Viswanathan, E. a Yates, Z. Shriver, A. Naggi, G. Torri, R. Sasisekharan, and M. Guerrini, "Insights into the human glycan receptor conformation of 1918 pandemic hemagglutinin-glycan complexes derived from nuclear magnetic resonance and molecular dynamics studies.," *Biochemistry*, vol. 53, no. 25, pp. 4122–35, Jul. 2014.

[6] G. L. Sassaki, S. Elli, T. R. Rudd, E. Macchi, E. A. Yates, A. Naggi, Z. Shriver, R. Raman, R. Sasisekharan, G. Torri, and M. Guerrini, "Human (alfa2->6) and avian (alfa2->3) sialylated receptors of influenza A virus show distinct conformations and dynamics in solution," *Biochemistry*, vol. 52, no. 41, pp. 7217–7230, 2013.

[7] Y. Chen, W. Liang, S. Yang, N. Wu, H. Gao, J. Sheng, H. Yao, J. Wo, Q. Fang, D. Cui, Y. Li, X. Yao, Y. Zhang, H. Wu, S. Zheng, H. Diao, S. Xia, Y. Zhang, K.-H. Chan, H.-W. Tsoi, J. L.-L. Teng, W. Song, P. Wang, S.-Y. Lau, M. Zheng, J. F.-W. Chan, K. K.-W. To, H. Chen, L. Li, and K.-Y. Yuen, "Human infections with the emerging avian influenza A H7N9 virus from wet market poultry: clinical analysis and characterisation of viral genome.," *Lancet*, vol. 381, no. 9881, pp. 1916–25, Jun. 2013.

[8] R. Gao, B. Cao, Y. Hu, Z. Feng, D. Wang, W. Hu, J. Chen, Z. Jie, H. Qiu, K. Xu, X. Xu, H. Lu, W. Zhu, Z. Gao, N. Xiang, Y. Shen, Z. He, Y. Gu, Z. Zhang, Y. Yang, X. Zhao, L. Zhou, X. Li, S. Zou, Y. Zhang, X. Li, L. Yang, J. Guo, J. Dong, Q. Li, L. Dong, Y. Zhu, T. Bai, S. Wang, P. Hao, W. Yang, Y. Zhang, J. Han, H. Yu, D. Li, G. F. Gao, G. Wu, Y. Wang, Z. Yuan, and Y. Shu, "Human infection with a novel avian-origin influenza A (H7N9) virus.," *N. Engl. J. Med.*, vol. 368, no. 20, pp. 1888–97, May 2013.

[9] K. Tharakaraman, A. Jayaraman, R. Raman, K. Viswanathan, N. W. Stebbins, D. Johnson, Z. Shriver, V. Sasisekharan, and R. Sasisekharan, "Glycan receptor binding of the influenza A virus H7N9 hemagglutinin.," *Cell*, vol. 153, no. 7, pp. 1486–93, Jun. 2013.

[10] T. Watanabe, M. Kiso, S. Fukuyama, N. Nakajima, M. Imai, S. Yamada, S. Murakami, S. Yamayoshi, K. Iwatsuki-Horimoto, Y. Sakoda, E. Takashita, R. McBride, T. Noda, M. Hatta, H. Imai, D. Zhao, N. Kishida, M. Shirakura, R. P. de Vries, S. Shichinohe, M. Okamatsu, T. Tamura, Y. Tomita, N. Fujimoto, K. Goto, H. Katsura, E. Kawakami, I. Ishikawa, S. Watanabe, M. Ito, Y. Sakai-Tagawa, Y. Sugita, R. Uraki, R. Yamaji, A. J. Eisfeld, G. Zhong, S. Fan, J. Ping, E. a Maher, A. Hanson, Y. Uchida, T. Saito, M. Ozawa, G. Neumann, H. Kida, T. Odagiri, J. C. Paulson, H. Hasegawa, M. Tashiro, and Y. Kawaoka, "Characterization of H7N9 influenza A viruses isolated from humans.," *Nature*, vol. 501, no. 7468, pp. 551–5, Sep. 2013.

[11] J. Zhou, D. Wang, R. Gao, B. Zhao, J. Song, X. Qi, Y. Zhang, Y. Shi, L. Yang, W. Zhu, T. Bai, K. Qin, Y. Lan, S. Zou, J. Guo, J. Dong, L. Dong, Y. Zhang, H. Wei, X. Li, J. Lu, L. Liu, X. Zhao, X. Li, W. Huang, L. Wen, H. Bo, L. Xin, Y. Chen, C. Xu, Y. Pei, Y. Yang, X. Zhang, S. Wang, Z. Feng, J. Han, W. Yang, G. F. Gao, G. Wu, D. Li, Y. Wang, and Y. Shu, "Biological features of novel avian influenza A (H7N9) virus.," *Nature*, vol. 499, no. 7459, pp. 500–3, Jul. 2013.

[12] X. Xiong, S. R. Martin, L. F. Haire, S. a Wharton, R. S. Daniels, M. S. Bennett, J. W. McCauley, P. J. Collins, P. a Walker, J. J. Skehel, and S. J. Gamblin, "Receptor binding by an H7N9 influenza virus from humans.," *Nature*, vol. 499, no. 7459, pp. 496–9, Jul. 2013.

ABBREVIATIONS

AV18: Double mutant (Asp225-Gly/Asp190-Glu) of SC18

dm: Double mutant

Glc: Glucose

Gal: Galactose

GlcNAc: N-acetylglucosamine

HA: Hemagglutinin

LSTa: Neu5Ac-α(2→3)-Gal-β(1→3)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc

LSTc: Neu5Ac-α(2→6)-Gal-β(1→4)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc

MD: Molecular Dynamic

NA: Neuraminidase

Neu5Ac: N-Acetylneuraminic acid or sialic acid

NMR: Nuclear Magnetic Resonance

NOE: Nuclear Overhauser Effect

NY18: Single mutant (Asp225-Gly) of SC18

PCA: Principal Component Analysis

RBS: Receptor Binding Site

SC18: Hemagglutinin from virus A/South Carolina/1/1918

sm: Single mutant

STD: Saturation Transfer Difference

PAPERS

1. Rudd T.R., **Macchi E.**, Gardini C., Muzi L., Guerrini M., Yates E.A., Torri G. "How To Find a Needle (or Anything Else) in a Haystack: Two-Dimensional Correlation Spectroscopy-Filtering with Iterative Random Sampling Applied to Pharmaceutical Heparin", Anal. Chem., **2012**, 84(15), 6841-6847.
2. Sassaki G.Ll, Elli S., Rudd T.R., **Macchi *E.***, Yates E.A., Naggi A., Shriver Z., Raman R., Sasisekharan R., Torri G., Guerrini M**.** "Human (α2→6) and Avian (α2→3) Sialylated Receptors of Influenza A Virus Show Distinct Conformations and Dynamics in Solution", *Biochemistry**, 2013***, 52, 7217−7230 .
3. Rudd T. R., **Macchi E.,** Muzi L., Ferro M., Gaudesi D., Torri G., Casu B., Guerrini M., Yates E.A. "Unravelling Structural Information from Complex Mixtures Utilizing Correlation Spectroscopy Applied to HSQC Spectra", *Anal. Chem.*, **2013**, *85* (15), 7487–7493.
4. Elli S.*, **Macchi E**.*, Rudd T. R.*, Raman R., Sassaki G.L., Viswanathan K., Yates E.A., Shriver Z., Naggi A., Torri G., Sasisekharan R., Guerrini M. "Insights into the human glycan receptor conformation of 1918 pandemic hemagglutinin-glycan complexes derived from nuclear magnetic resonance and molecular dynamics studies." *Biochemistry*, **2014**, *53*(25), 4122–35.
5. Coletti A., Elli S., **Macchi E**., Galzerano P., Zamani L., Guerrini M.,Torri G., Vismara E. "Conformational changes of 1-4-glucopyranosyl residues of a sulfated C-C linked hexasaccharide." *Carbohydrate Research*, **2014**, *389*, 134–40.

ORAL COMUNICATIONS

1. E. Macchi, S. Elli, T. R. Rudd, R. Raman, G. Sassaki, K. Viswanathan, E. Yates, Z. Shriver, A. Naggi, G. Torri, R. Sasisekharan, and M. Guerrini *"NMR And MD Investigation Of The 1918 Pandemic Heamagglutinin-Glycan Receptor Complex"* XiV Convegno Scuola sulla Chimica dei Carboidrati, 22-25 Giugno 2014, Pontignano (Siena).
2. S. Elli, E. Macchi, T. R. Rudd, R. Raman, G. Sassaki, K. Viswanathan, E.A. Yates, Z. Shriver, A. Naggi, G. Torri, R. Sasisekharan, and M. Guerrini *"MD and NMR investigation of 1918 pandemic influenza virus haemagglutinin-glycan receptor complexes. Novel insights into human/avian influenza virus recognition specificity."* Workshop Bioactive carbohydrate:syntetic, structural and analytical aspects, Milano, 18 Dicembre 2013
3. S. Elli, E. Macchi, T. R. Rudd, R. Raman, G. Sassaki, K. Viswanathan, E.A. Yates, Z. Shriver, A. Naggi, G. Torri, R. Sasisekharan, and M. Guerrini *"Conformational and dynamic properties of Human (α2->6) and Avian (α2->3) sialylated receptors of influenza A virus haemagglutinin"*, 2nd EAST-NMR meeting, Bratislava, 2012

COMUNICATIONS

1. Sassaki G., Elli S., Rudd TR., Macchi E., Yates E.A., Naggi A., Shriver Z., Raman R., Sasisekharan R., Torri G. and Guerrini M. "Human (α2→6) and avian (α2→3) sialylated receptors of influenza A virus show distinct conformations and dynamic in solution". Breakthroughs in NMR of Structural Biology – 2nd Bio-NMR Annual User Meeting, Portorož, Slovenia (Poster)
2. SassakiL. G., Elli S., Rudd T. R., Macchi E., Raman R., Shriver Z., Sasisekharan R., Torri G., Guerrini M. "Distinct topologies of human-like (α2-6 sialylated) and avian-like (α2-3 sialylated) glycan receptors govern their binding to haemagglutinin from influenza viruses". XIII Convegno-Scuola sulla Chimica dei Carboidrati, Pontignano (Si) (Poster)

3. Macchi E., Elli S., Rudd T. R., Sassaki G., Shriver Z., Raman R., Sasisekharan R., Torri G., Guerrini M. "Selected amino acids mutation of Human and Avian Influenza A hemagglutinin affect the binding to the upper respiratory epithelia glycan receptors: NMR and MD structural investigation", *NMR and protein dynamics in structural biology- The 3rd Bio-NMR Annual User Meeting, Budapest, Hungary 10-13 June 2013 (Poster)*
4. Macchi E., Elli S., Rudd T., Shriver Z., Raman R., Sasiskharan R., Torri G., Guerrini M. " NMR and MD studies of the new H7N9 hemagglutinin with human and avian glycosidic receptors, The 4th Bio-NMR User meeting, *May 2014, Varsaw (Poster)*

RINGRAZIAMENTI

"Insegnami la dolcezza ispirandomi la carità, insegnami la disciplina dandomi la pazienza e insegnami la scienza illuminandomi la mente."

Ed eccomi qui, mentre il sipario si sta per chiudere su un altro "atto" della mia vita, mi sembra doveroso ringraziare tutte le persone che l'hanno reso possibile.

Il primo grazie va sicuramente a mamma e papà, perchè senza di loro nulla sarebbe stato possibile, perchè hanno sempre creduto in me, soprattutto quando io stessa non riuscivo a farlo, perchè quello che sono lo devo soprattutto a loro e per questo non smetterò mai di ringraziarli.

Il secondo grazie va a Stefano, perchè mi è vicino sempre e comunque, anche quando la tensione è alta e lo "sclero facile" in agguato. Perchè sei di una bontà infinita, che spesso spiazza, ma sai anche essere fermo e testone a dovere per tenermi testa, per frenarmi e per "mettermi al mio posto" quando ce n'è bisogno.

Il terzo grazie va sicuramente a Eli, la mia twin, perchè è la mia twin! Perchè se non ci fosse bisognerebbe inventarla! Perchè anche se siamo due gemelle "molto diverse" sono contenta di averti sempre al mio fianco!

Non posso certo dimenticare Sissa, che dall'avventura della tesi specialistica…non mi ha più mollato! Perchè mi hai aiutato a crescere e perchè con te so di avere sempre una spalla su cui appoggiarmi!

Le mie socie, Eli e Coppi..che dire..:GRAZIE! Perchè anche se ora siamo lontane, ci siete sempre e comunque! Perchè basta un ristorante cinese ed è come se questi quattro e più anni non fossero mai passati! Perchè sono orgogliosa e felice di avervi come AMICHE!

Come dimenticare poi le persone che con il loro lavoro hanno reso possibile il mio: tutto lo staff del Ronzoni, senza dimenticare nessuno! Un grazie di cuore a tutti!

Un ringraziamento particolare va però al gruppo NMR! Marco Guerrini (il boss) e Stefano Elli (MD man) perchè senza di loro questo traguardo non sarebbe stato possibile! Poi Laura, per i consigli, le chiacchierate e gli sfoghi! Cesare e Lucio per il supporto morale! E poi Tim, per le correzioni al mio inglese e a tutti i suoi preziosi consigli e suggestion anche da oltre manica!

Infine un grazie ad Hans per il support tecnico e morale ad Utrecht sul 900 MHz, a Giacomo, Davide e Francesca per il supporto vario ed eventuale in questi anni!

Annex

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Human (α2→6) and Avian (α2→3) Sialylated Receptors of Influenza A Virus Show Distinct Conformations and Dynamics in Solution

Guilherme L. Sassaki,[†,∥] Stefano Elli,[†] Timothy R. Rudd,[†,⊥] Eleonora Macchi,[†] Edwin A. Yates,[‡] Annamaria Naggi,[†] Zachary Shriver,[§] Rahul Raman,[§] R. Sasisekharan,[§] Giangiacomo Torri,[†] and Marco Guerrini*[,†]

[†]Istituto di Ricerche Chimiche e Biochimiche "G. Ronzoni", Via Giuseppe Colombo, 81, Milano, 20133 Italy

[‡]Department of Structural and Chemical Biology, Biosciences Building, University of Liverpool, Crown Street, Liverpool L69 7ZB, U.K.

[§]Harvard-MIT Division of Health Sciences and Technology, Koch Institute for Integrative Cancer Research, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

Supporting Information

ABSTRACT: Differential interactions between influenza A virus protein hemagglutinin (HA) and α2→3 (avian) or α2→6 (human) sialylated glycan receptors play an important role in governing host specificity and adaptation of the virus. Previous analysis of HA−glycan interactions with trisaccharides showed that, in addition to the terminal sialic acid linkage, the conformation and topology of the glycans, while they are bound to HA, are key factors in regulating these interactions. Here, the solution conformation and dynamics of two representative avian and human glycan pentasaccharide receptors [LSTa, Neu5Ac-α(2→3)-Gal-β(1→3)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc; LSTc, (Neu5Ac-α(2→6)-Gal-β(1→4)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc] have been explored using nuclear magnetic resonance and molecular dynamics simulation. Analyses demonstrate that, in solution, human and avian receptors sample distinct conformations, topologies, and dynamics. These unique features of avian and human receptors in solution could represent distinct molecular characteristics for recognition by HA, thereby providing the HA−glycan interaction specificity in influenza.

The interaction between the influenza A virus surface protein hemagglutinin (HA) and its cell surface glycan receptors is one of the critical steps governing virus selectivity. Several crystal structures of different HAs have been determined, both native and cocrystallized with glycan receptors.[1−5] These studies have provided insight into the ability of HA to interact with receptor glycans. Human-adapted influenza viruses preferentially bind to sialylated glycan receptors terminated by *N*-acetylneuraminic acid (Neu5Ac) α2→6 linked to galactose (Gal) (human receptors), which are predominantly expressed on the apical surface of the upper respiratory epithelia in humans.[6−8] In contrast, avian-adapted influenza viruses preferentially bind to sialylated glycans terminated by Neu5Ac α2→3 linked to Gal, which are expressed predominantly in the gut and respiratory tract of birds and, importantly, in the deep lung (alveolar region) of humans.[7] Therefore, one of the characteristics of human adaptation of influenza A viruses is the change in glycan receptor binding preference of HA from avian to human receptors.[9−14]

Recent studies based on analysis of HA−glycan cocrystal structures[15,16] have highlighted the importance of glycan conformation and topology in governing the receptor specificity of avian- and human-adapted HAs. In the case of avian receptors, the majority of contacts between HA and glycan involve the Neu5Ac residue at the nonreducing end. Chandrasekaran et al.[16] proposed that, when the Neu5Ac is fixed in the glycan-binding pocket of HA, the other residues in this trisaccharide motif sample different torsional angles at their glycosidic linkages, such that the space occupied by the whole glycan resembles a cone with Neu5Ac at the vertex, characterized by a topological angle θ of >110°. The topological angle θ is defined by the C2, C1, and C1 atoms of the residues Neu5Ac, Gal, and GlcNAc (*N*-acetylglucosamine), respectively, in going from the nonreducing end to the reducing end. They also proposed for the case of human receptors that a longer glycan motif, comprising at least a tetrasaccharide Neu5Ac-α(2→6)-Gal-β(1→4)-GlcNAc-β(1→3)-Gal- at the nonreducing end, spans the glycan binding surface on HA. The additional C5−C6 bond of the Neu5Ac-α(2→6)-Gal glycosidic linkage provides added mobility to its glycosidic junction and the remaining backbone, allowing it to explore conformations

Received: May 29, 2013
Revised: September 9, 2013
Published: September 9, 2013

ACS Publications © 2013 American Chemical Society

similar to an umbrella progressing from a fully open form to a fully closed form (defined as having a θ of <110°).[16] It is clear that the cone and umbrella topologies are generated by their motions while they are bound to the receptor, but when in their free states, they adopt conformations that are dependent on their respective nonreducing end connectivity, $\alpha(2\rightarrow3)$ or $\alpha(2\rightarrow6)$. The role of glycan shape and topology in regulating binding specificity for HA has also been investigated using molecular dynamics simulations,[17] indicating that there are distinct differences in the conformational sampling of avian and human receptors from the different subtypes (H3, H5, and H9) upon binding to HA.

The previously mentioned studies highlighted differences in glycan receptor conformation and structural details when the receptor cocrystallized bound to HA, illustrating the structural specificity of HA−glycan interactions. Consequently, it is reasonable to suppose that receptor dynamic properties, strictly related to its structure, could play a role in the recognition and binding by the HA protein. To characterize human and avian sialylated glycan receptors in terms of conformation and topology in solution, detailed nuclear magnetic resonance (NMR) analyses, MD simulations, and subsequent numerical analysis of the data generated were performed on the two sialylated pentasaccharides: LSTc (Neu5Ac-$\alpha(2\rightarrow6)$-Gal-$\beta(1\rightarrow4)$-GlcNAc-$\beta(1\rightarrow3)$-Gal-$\beta(1\rightarrow4)$-Glc), representing the human receptor, and LSTa (Neu5Ac-$\alpha(2\rightarrow3)$-Gal-$\beta(1\rightarrow3)$-GlcNAc-$\beta(1\rightarrow3)$-Gal-$\beta(1\rightarrow4)$-Glc), representing the avian receptor (Figure 1 of the Supporting Information). We refer to Gal1 and Gal2 as labels for the first and second Gal residues encountered in the sequence of LSTc and LSTa, respectively, from the nonreducing to the reducing end. NOESY, ^{13}C NMR relaxation measurements, and analysis of the glycan MD trajectories revealed that LSTa and LSTc have distinct conformational dynamics, topologies, and intramolecular hydrogen bond networks in solution. Specifically, LSTc preferentially adopts a bent shape in solution, which is consistent with the "umbrella"-like conformation, whereas LSTa adopts a bow shape in solution, which agrees with a "cone"-like structure. These observations suggest that the different species-adapted HAs could recognize specific shapes, conformations, and dynamic behaviors of glycan in solution.

■ EXPERIMENTAL PROCEDURES

NMR Chemical Shift Mapping. NMR measurements were taken using Bruker 500, 600, and 900 MHz AVANCE series NMR spectrometers (Bruker GmbH, Silberstreifen, Germany) with a high-sensitivity 5 mm TCI cryoprobe. Human milk lactosyl sialyl-*N*-tetraoses, namely, LSTa (PROzyme, Hayward, CA) [Neu5Ac-$\alpha(2\rightarrow3)$-Gal(Gal1)-$\beta(1\rightarrow3)$-GlcNAc-$\beta(1\rightarrow3)$-Gal(Gal2)-$\beta(1\rightarrow4)$-Glc)] and LSTc (PROzyme) [Neu5Ac-$\alpha(2\rightarrow6)$-Gal(Gal1)-$\beta(1\rightarrow4)$-GlcNAc-$\beta(1\rightarrow3)$-Gal(Gal2)-$\beta(1\rightarrow4)$-Glc], were dissolved in D_2O (99.9%) or a buffered solution [150 mM sodium chloride, 100 mM sodium phosphate, and 10% D_2O (pH 7.2)] according to the NMR experiment. Chemical shifts (^{1}H and ^{13}C) of the oligosaccharides were measured at 295 K in D_2O (99.9%, 2 mg/0.2 mL) using a 3 mm NMR tube. Two-dimensional (2D) NMR experiments were conducted using edited HSQC, HSQC-TOCSY, HMBC, TOCSY, and NOESY (using different mixing times between 100 and 1000 ms), recorded for quadrature detection in the indirect dimension and acquired using 16−48 scans per series of 1024 × 320 data points, with zero filling in F1 (4096) prior to Fourier transformation. Figures of spectra were created using R[18] and the libraries rNMR[19] and lattice.[20]

Estimation of Intra- and Interproton Distances. Experimental distances were obtained by selective one-dimensional (1D) and 2D NOESY experiments conducted with mixing times between 100 and 800 ms (295 K) and used to construct the kinetic NOEs. Selected distances were estimated on the basis of the (two-spin model approximation) time-dependent NOE measurements extrapolated to short mixing times, considering the initial buildup enhancement up to 300 ms.[21] The extrapolated NOEs signals were converted to proton distances considering the Neu5Ac distance ($H3_{ax}$−$H3_{eq}$) as a reference (1.78 Å).

^{13}C NMR Relaxation Data. NMR relaxation times were obtained at 295 K using a 14.0 T magnetic field. ^{13}C longitudinal relaxation (T_1) and transverse relaxation (T_2) were measured via 2D double INEPT-based inverse detection methods optimized to suppress the cross correlation between the chemical shift anisotropy (CSA) and the dipolar relaxation.[22−24] The 2D spectra were recorded using 1024 × 192 data points, recording 24−32 scans at eight different relaxation delays arranged randomly to avoid systematic errors in every experiment. T_1 relaxation delays were 0, 0.1, 0.2, 0.3, 0.5, 0.8, 1.0, and 1.4 s. T_2 transverse relaxation times were obtained by the CPMG pulse sequences differing in spin echo delays (22.4, 44.8, 67.2, 89.6, 134.4, 179.2, 224.0, and 313.6 ms). The volumes of the cross-peaks were integrated using Topspin 2.1, and relaxation times were obtained by best nonlinear and linear fitting of normalized data of the evolution of the peak volumes as a function of the inversion recovery delays. All the values and estimated errors (2−10%) were calculated and plotted in Origin 8.0 (OriginLab Corp., Northampton, MA).

Determination of the $^nJ_{C-H}$ Heteronuclear Coupling Constants. Long-range J_{C-H} couplings were measured using the J-HMBC pulse sequence (295 K), with a delta delay corresponding to a $^nJ_{X-H}$ of 2 Hz [LSTc (0.7 mg) and LSTa (1 mg) were dissolved in 220 μL of D_2O]. Data matrices of 1024 × 320 points covering 2003.2 × 9054.9 Hz were expanded to 2048 × 1024 points using linear prediction and subsequently zero-filled to 8192 × 2048 points prior to echo/antiecho Fourier transformation. The traditional sequence was applied for LSTa, while the non uniform sampling (NUS) variant was used for LSTc (Bruker Topspin version 3.1).

Determination of Temperature Coefficients and Chemical Shift Deviations. Temperature coefficients were recorded for 0.5 mg/0.2 mL oligosaccharide solutions in buffered D_2O or in a 90:10 H_2O/D_2O mixture at pH 7.2. All temperature gradients were measured by differences observed in spectra recorded at 3−9 K intervals between 278 and 313 K (^{1}H). Temperature coefficients are expressed in units of parts per billion per kelvin, with a negative sign indicating an upfield shift upon warming; chemical shifts were measured using ^{1}H and HSQC experiments.

Molecular Dynamics Simulations. The starting geometries of both oligosaccharides LSTa and LSTc were generated using the carbohydrate biomolecule builder;[25] following the primary structure and the glycosidic constraints, both structures were minimized by the online energy minimizer.[25] The energy description of both oligosaccharides included the amber/glycam06 force field.[25,26] The atomic partial charges required by the oligosaccharide force field were calculated using the semiempirical bond-corrected AM1-BCC

scheme.[27] The MD simulations were conducted in explicit solvent using the TIP3P water model.[28] Oligosaccharide partial charge management and topology files were constructed with AmberTools version 1.4.[29] The MD trajectories were generated using the NAMD MD engine.[30] Nonbonded interactions, i.e., electrostatic and van der Waals, are described using the cutoff approach set to 12.0 Å with a switching function beginning at 6.0 Å, and the time step for the integration algorithm is 1 fs. The simulation cell, including the oligosaccharide, was built enveloping the solute molecule in a layer of water 20.0 Å wide in each direction; the result is a cubic cell of approximately 60 Å × 67 Å × 59 Å. The oligosaccharide models that were compared to the experimental data were generated by the following procedure; the simulation cell was first optimized applying 10000 minimization steps of the NAMD minimizer algorithm, restraining each atom of the carbohydrate to the initial conformation by a harmonic restraint with a scaling constant of 500 kcal/mol. The next step was a MD simulation reproducing the constant-temperature and pressure statistical ensemble (*NPT*) applying a pressure on the cell walls of 1.01325 bar (atmospheric pressure). The simulation temperature was set at 295 K and was controlled by a Langevin thermostat as implemented in NAMD. The Nosé-Hoover Langevin piston algorithm controlled the pressure applied on the cell walls, with periodic boundary conditions being applied. The whole length of the simulation was 104 ns, where the first 4 ns was conducted restraining the carbohydrate atoms (solute) by a harmonic potential with a scaling constant of 50 kcal/mol to initially equilibrate the cell density without changing the solute conformation significantly, while the remaining simulation time was undertaken with the solute free to move. The cell densities calculated for LSTa and LSTc after equilibration (~4 ns) were in both cases 1.04 g/cm^3. The MD trajectory consisted of the last 100 ns, with a single geometry being sampled every 10 ps. During the whole dynamic, the conformation of the residues remained in $^2C_5$ for Neu5Ac and $^4C_1$ for the others, as set from the beginning for both LSTa and LSTc, and in agreement with the three-bond proton−proton coupling constants ($^3J_{H-H}$).

The progression of the MD simulation was observed by determining the root-mean-square displacement (rmsd) of the trajectories, performed in R using the *rmsd* function from the Bio3D library,[31] with the starting conformation as the reference. The rmsd distances were determined using eq 1:

$$\mathrm{rmsd}(t) = \sqrt{n^{-1}\sum_{i=1}^{n}[D_i(t) - D^{\mathrm{ref}}{}_i]^2} \quad (1)$$

where $D^{\mathrm{ref}}{}_i$ is the position of the *i*th atom of the reference structure ($t = 0$), $D_i(t)$ is the position of atom *i* at time *t*, and *n* is the number of atoms.

Glycosidic Torsion Angles. The torsional angles (ϕ and ψ) are defined in this paper by the following pairs of dihedral angles, from the nonreducing termini: ϕ_1 and ψ_1, ϕ_2 and ψ_2, ϕ_3 and ψ_3, and ϕ_4 and ψ_4. For LSTa, ϕ_1 is the C1−C2−O3−C3 angle and ψ_1 is the C2−O3−C3−H3 angle, while for LSTc, ϕ_1 is the C1−C2−O6−C6 angle and ψ_1 is the C2−O6−C6−C5 angle, in agreement with Xu et al.[17] The successive pairs are defined as ϕ_i and ψ_i [H1−C1−O4′−C4′ and C1−O4′−C4′−H4′ (1→4 linkage) or H1−C1−O3′−C3′ and C1−O3′−C3′−H3′ (1→3 linkage)], as conventionally defined for sugars. Atoms labeled with a prime belong to the monosaccharide on the reducing side of the glycosidic linkage, while atoms without a prime are on the nonreducing side of the glycosidic linkage. To be consistent with the work of Chandrasekaran et al.[16] and Xu et al.,[17] ϕ and ψ were illustrated in the range of −120° to 240°.

Topological θ Angle. The θ angle is defined by the C2, C1, and C1 atoms of residues Neu5Ac, Gal, and GlcNAc (*N*-acetylglucosamine), respectively, going from the nonreducing end to the reducing end

Comparison of the MD Results to the Experimental Data. The $\mathrm{rmsd}_{\mathrm{exp}}$ values (eq 2) were used to compare the MD trajectories with the experimental data, i.e., the estimated inter-residue proton−proton distances for both LSTc and LSTa using the two-spin model approximation, reported for LSTc (H5 Neu5Ac−CH_3 GlcNAc, H1 GlcNAc−H3 Gal2, H1 GlcNAc−H4 Gal2, H1 Gal2−H4 Glc, and H1 Gal1−H4 GlcNAc) and LSTa ($H3_{ax}$ Neu5Ac−H3 Gal1, $H3_{eq}$ Neu5Ac−H3 Gal1, H1 Gal1−H3 GlcNAc, H1 Gal1−CH_3 GlcNac, H1 GlcNAc−H3 Gal2, H1 GlcNAc−H4 Gal2, H1 Gal2−H4 Glc, and H1 Gal2−H5 Glc). In eq 2, $D_{\mathrm{exp},i}$ is the experimental *i*th inter-residue proton−proton distance and $D_{\mathrm{MD},i}$ is the corresponding distance for a single MD-generated conformation. The number of the available inter-residue proton−proton distances is *n*.

$$\mathrm{rmsd}_{\mathrm{exp}} = \sqrt{n^{-1}\sum_{i=1}^{n}(D_{\mathrm{exp},i} - D_{\mathrm{MD},i})^2} \quad (2)$$

2D NOESY Spectral Simulation (multispin model relaxation). The 2D NOESY spectra were simulated using two different sets of 40 conformations each from the MD trajectory: 40 with the smallest rmsd ($\mathrm{rmsd}_{\mathrm{exp}}$ min) and highest ($\mathrm{rmsd}_{\mathrm{exp}}$ max). The 2D NOESY simulation was undertaken assuming isotropic motion with a single correlation time τ_c estimated to be 400 ps. The τ_c estimation was conducted on the smallest rmsd set of conformations ($\mathrm{rmsd}_{\mathrm{exp}}$ min) and simulating the intraresidue H1−H3 and H1−H5 2D NOE signals for the selected residues: GlcNAc for LSTc and Gal1, GlcNAc, and Gal2 for LSTa. The τ_c parameter was adjusted by trial and error until the NOE kinetic enhancement from 0 to 500 ms mixing time reproduced qualitatively the corresponding experimental values. The software for 2D NOESY simulation was NOEPROM.[32]

Principal Component Analysis of the MD Trajectory. The MD trajectories for LSTc and LSTa were exported as *XYZ* coordinates via VMD,[33] allowing the data to be imported directly into R.[18] Principal component analysis (PCA) was performed using the bio3d library, the *pca.xyz* function.[31] The data were referred to the first conformer of the MD trajectory, removing translational and rotational effects from the analysis, and then mean centered before PCA was performed. The calculated eigenvalues (scree plot) provide a weighting for the importance of the eigenvectors (also known as loadings) (illustrated as porcupine plots), and the component scores (score plot) illustrate the relationship between the determined eigenvector (independent molecular motion) and the conformers that constitute the MD trajectory. Porcupine plots were produced by constructing a trajectory along the desired principal component, imposing the molecular motion extracted by that component onto the average geometry of the conformers.[31] This was then visualized using VMD and the porcupine plot created using a VMD script.[33,34]

The convergence of the glycan PCA was observed by determining the root-mean-square internal product (rmsip) of

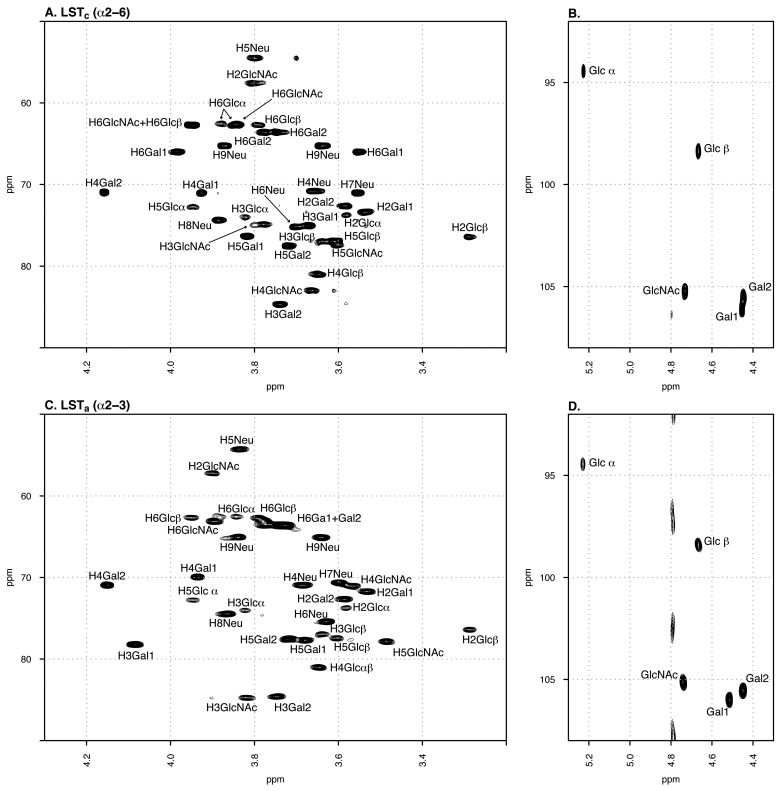


Figure 1. $^{1}H-^{13}C$ HSQC NMR spectra of LSTc (A and B) and LSTa (C and D).

the first 10 eigenvectors. If rmsip is equal to 1, then there is complete convergence between the sets of eigenvectors; a value of ≥0.70 shows that the eigenvectors are consistent with each other.[35] In this circumstance, rmsip was used to see if the MD simulation was tending toward a stationary state. PCA was performed on the trajectory at different times, and this was compared to the eigenvectors derived for the trajectory at 100 ns. The rmsip was determined using R and the library Bio3d;[31] the *rmsip* function was utilized. The rmsip is defined as

$$\text{rmsip} = \sqrt{10^{-1}\sum_{i=1}^{10}\sum_{j=1}^{10}(\eta_i \cdot \nu_j)^2} \quad (3)$$

where η_i is the ith eigenvector of trajectory η and ν_i is the ith eigenvector of trajectory ν.

RESULTS

NMR Analyses of LSTa and LSTc. NMR characterization of sialyl-lacto-*N*-tetraoses LSTa and LSTc (Figure 1 of the Supporting Information) were conducted using homonuclear (COSY, TOCSY, and NOESY), heteronuclear (edited HSQC and HSQC-TOCSY), and heteronuclear multiple-bond correlation HMBC ($^{1}H/^{13}C$) experiments. This permitted the complete assignment of the hydrogen and carbon resonances for both LSTc and LSTa (Figure 1 and Table 1 of the Supporting Information). TOCSY analysis allowed the

connectivity of the pyranosyl rings to be observed, with H−H axial correlations confirming the monosaccharide identities. The chemical shifts of the acetamide methyl groups of GlcNAc and Neu5Ac for both LSTa and LSTc were determined by 2D TOCSY (buffered system, 10% D_2O), providing NH correlations with some of the hydrogens of each ring and with those from acetamide (Figure 2 of the Supporting Information). However, the strong peak superimposition did not permit the complete hydrogen assignment by homonuclear experiments; these were complemented by HSQC-TOCSY and by the COSY effect generated by homonuclear coupling in sensitivity-enhanced HSQC. These correlation peaks, usually considered as artifacts, are very useful in the assignment of signals in crowded spin systems when performing sensitivity-enhanced HSQC experiments[36] (Figure 3 of the Supporting Information). Long-range correlations ($^1H/^{13}C$) of the quaternary C1 and C2 atoms from the sialyl residues were observed between the $H3_{ax}$ Neu5Ac atom at 1.782 ppm and the corresponding carbon signals at 176.7 and 102.3 ppm (C1 and C2, respectively). C2 correlated with H3 Gal1 at 4.082 and 78.2 ppm of LSTa, confirming the presence of the Neu5Ac $\alpha2\rightarrow3$ Gal linkage. Instead, LSTc correlated $H3_{ax}$ Neu5Ac at 1.711 ppm with carbon signals at 176.2 (C1) and 102.7 (C2) ppm, with the latter correlating with H6α/β Gal1 at 3.983α/3.S44β/66.0 ppm, confirming that C2 Neu5Ac is O-linked to H6α/β Gal1. Anomeric proton correlations from LSTa were observed at 4.513/84.7 ppm (H1 Gal1/C3 GlcNAc), 4.739/84.5 ppm (H1 GlcNAc/C3 Gal2), and 4.446/81 ppm (H1 Gal2/C4 Glc). LSTc exhibited cross-peaks at 4.451/82.9 ppm (H1 Gal1/C4 GlcNAc), 4.73/84.6 ppm (H1 GlcNAc/C3 Gal2), and 4.445/80.9 ppm (H1 Gal2/C4 Glc). These measurements were confirmed by 1D selective and 2D NOESY experiments.

Three-bond proton−proton coupling constants, $^3J_{H-H}$, were measured for both pentasaccharides, which indicated that all Hex*p* residues assumed the $^4C_1$ conformation, because $^3J_{H1-H2}$ values vary from 7.9 to 8.2 Hz and $^3J_{H3-H4}$ values for Gal residues are ∼3.5 Hz; Glc and GlcNAc had values of 9.0 and 9.8 Hz, respectively. The Neu5Ac residue had $^3J_{H-H}$ values compatible with those of a pyranosyl ring in the $^2C_5$ conformation [$^3J_{H3-H4}$ = $4.8_{(eq)}$ Hz/$12.1_{(ax)}$ Hz, and $^3J_{H4-H5}$ = 10.1 Hz], thus suggesting no significant differences in the ring conformation between the pentasaccharides.

As in proteins, NH temperature coefficients can reflect the extent to which *N*-acetyl-amide protons are protected from exchange through hydrogen bonding. Evaluation of NH temperature coefficients showed different behavior between GlcNAc and Neu5Ac. The first had a value of −8.28 ppb/K for LSTc and −8.00 ppb/K for LSTa, while the NH group for Neu5Ac exhibited lower temperature coefficients with values of −6.77 and −6.60 ppb/K for LSTc and LSTa, respectively. This behavior suggested that the amide group from GlcNAc could be involved in an intramolecular hydrogen bond. This result is supported by a study of the half-height line widths ($\Delta\nu_{1/2}$) extracted from each NH acetamide after a Lorentzian deconvolution and plotted at different temperatures. The NH data from the Neu5Ac *N*-acetyl group broadened before those of GlcNAc, suggesting that the NH group of GlcNAc is involved in intramolecular hydrogen bonds (Figure 4 of the Supporting Information).

^{13}C NMR Relaxation Data. NMR relaxation studies have been utilized to interrogate the internal motion of the glycans, providing dynamic properties that can be qualitatively or quantitatively assessed.[23,24,37,38] T_1 and T_2 relaxation times and heteronuclear NOEs were measured at 14.0 T (Table 2 of the Supporting Information). T_1 varied from 315 to 412 ms for LSTa and from 328 to 421 ms for LSTc. T_2 varied from 229 to 355 ms and from 232 to 387 ms for LSTa and LSTc, respectively. To qualitatively observe the main differences in the motional properties between LSTc and LSTa, the average values of T_1 and T_2 of each CH ring were used (Figure 2). The

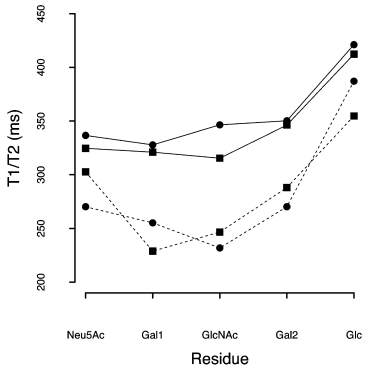


Figure 2. Average ^{13}C relaxation times of CH group carbons for each monosaccharide unit from LSTc (●) and LSTa (■). The solid line represents the average T_1 of each monosaccharide, while the dashed line represents the average T_2 of each monosaccharide.

numerical values of T_1 for the different residues suggest an anisotropically tumbling molecule, because the Neu5Ac-Gal1-GlcNAc trisaccharide of both glycans gave rise to T_1 values ranging from 315 to 346 ms, whereas values of the reducing Glc residue varied from 410 to 421 ms. This finding indicates a higher degree of motion for the reducing lactosyl disaccharide than for the nonreducing end of the oligosaccharide. The heteronuclear NOEs corroborate the T_1 values, because larger heteronuclear NOEs are observed for the terminal units. Using the same approach, significantly lower T_2 values were detected for both LSTa and LSTc, T_2 values of the reducing Glc residue of both glycans (355 and 387 ms for LSTc and LSTa, respectively) confirmed the higher degree of motion of this residue than of the internal Gal1, and GlcNAc residues showed lower T_2 values between 228 and 255 ms for both glycans. The main differences between LSTa and LSTc were the values observed for T_2 of the Neu5Ac residues, being 303 and 270 ms, respectively; such a difference, larger than the experimental error, accounts for the significantly higher rigidity of the nonreducing end of LSTc. Contrary to that of the nonreducing side, the higher mobility of the reducing end residues has little biological relevance, because this part of the molecule *in vivo* is part of a greater structure.

Inter- and Intra-NOESY Experimental Distances. The inter- and intraresidue proton−proton distances were measured by NOESY enhancement (see Experimental Procedures for details). Intraresidue distances (H1−H3 of Hex*p* units and H3−H5 of the Neu5Ac residues) were measured between 2.0

and 2.6 Å for all the residues (Table 3 of the Supporting Information). These values were in agreement with the expected $^4C_1$ and $^2C_5$ ring chair conformations, supporting the observed $^3J_{H-H}$ coupling constants. Inter-residue proton−proton distances were obtained mainly from O-glycosidic bonds (Table 1). LSTa exhibited NOE $H3_{eq}$ and $H3_{ax}$ Neu−H3

Table 1. Selected Interglycosidic Proton−Proton Distances (angstroms) for LSTa and LSTc Derived from NOE Measurements

	LSTc (Å)	LSTa (Å)
$H3_{ax}$ Neu5Ac−H3 Gal1	–	2.76
$H3_{eq}$ Neu5Ac−H3 Gal1	–	4.70
H5 Neu5Ac−CH_3 GlcNAc	4.24	–
H1 Gal1−H3 GlcNAc	–	2.25
H1 Gal1−CH_3 GlcNAc	–	4.00
H1 GlcNAc−H3 Gal2	1.94	1.95
H1 GlcNAc−H4 Gal2	3.05	2.84
H1 Gal2−H4 Glc	1.96	2.20
H1 Gal2−H5 Glc	–	2.80

Gal1 contacts, providing distances of 4.7 and 2.8 Å, respectively, while LSTc exhibited an inter-residue NOE between H5 Neu5Ac and CH_3 GlcNAc corresponding to an estimated distance of 4.2 Å (Figure 5 of the Supporting Information). This contact between GlcNAc and Neu5Ac residues, observed only in the NOESY spectra of LSTc, can be considered the experimental proof of a bent conformation that LSTc assumes compared to LSTa, which does not exhibit this signal at all.

Molecular Dynamics Simulations of LSTa and LSTc. To complement the NMR experiments, MD simulations were performed to compare the dynamics of the glycans in solution. The MD simulations were allowed to run for 100 ns, sampling conformers every 0.01 ns, therefore producing a comprehensive data set, with the simulation length considerably greater than the estimated correlation time (isotropic motion) of either glycan, of ∼400 ps. After 70% of the simulation time had elapsed, several molecular parameters exhibited stationary behavior without significant drift; these include the end-to-end distance of the sugars, which is defined as the distance between C5 of Neu5Ac (nonreducing end) and C1 of Glc (reducing end). This property is one of the slowest-evolving degrees of freedom for an oligomer (Figures 7 and 8 of the Supporting Information). Stationary behavior is also observed for the θ angle of the glycans, which is characteristic for each glycan (Figures 9 and 10 of the Supporting Information) and the ω torsional angle of Gal1, which is defined as the H5−C5−C6−O6 angle (Figure 11 of the Supporting Information). Similar behavior was seen over the same time scale for the other four pairs of glycosidic torsional angles (ϕ_i and ψ_i) (see Experimental Procedures), the angles that define the backbone conformation of the two oligomers.

During the MD simulation, the residues comprising LSTc and LSTa adopted the $^2C_5$ (Neu5Ac) and $^4C_1$ (Gal, Glc, and GlcNAc) conformations, respectively, with only small deviations from the ideal chair of cyclohexane; the intraresidue proton−proton distances (H1−H3, H1−H5, and H3−H5) have values ranging between 2.5 and 2.7 Å for the five sugars, compared to the ideal value of 2.64 Å for cyclohexane. The same distances are estimated by NOE enhancement using the two-spin approximation (see Experimental Procedures); the values obtained ranged between 2.0 and 2.6 Å, in good agreement considering the strong coupling effects (Table 3 of the Supporting Information).

The intersugar unit conformations were defined using four pairs of inter-residue glycosidic dihedral angles, running from the nonreducing end to the reducing end of the glycans (ϕ_i and ψ_i) (see Experimental Procedures). The torsional angle values explored during the whole MD simulation are reported in Figure 3, and the torsional angles are displayed in the range of −120 to 240°. Visited torsional states are reported as cluster centroids with their relative populations in Table 2. The principal difference between LSTa and LSTc is located at the first glycosidic linkage (ϕ_1 and ψ_1), which is in agreement with those obtained by molecular modeling,[39] and is related to the different connectivity at the nonreducing end [Neu5Ac-α(2→3)-Gal for LSTa and Neu5Ac-α(2→6)-Gal for LSTc]. For LSTa, four ϕ_1 and ψ_1 torsional states are observed, a principal state is located at −62° and −8° (89.8%), respectively, while the most significant member of the minor states is located at −83° and −55° (7.7%), respectively (Figure 3A, α and β, respectively). LSTc possesses two ϕ_1 and ψ_1 torsional states; the cluster with the greater population (96.2%) is located at −57° and 190°, respectively, and the minor population at 56° and 174°, respectively (Figure 3E, ξ and o, respectively).

To fully describe the first glycosidic linkage of LSTc [Neu5Ac-α(2→6)-Gal], the ω torsional angle (see Experimental Procedures for the definition) needed to be considered. This degree of freedom is absent for LSTa. LSTc has a single state characterized by an average ω value of −50°, which is in agreement with the prediction of Xu et al.[17] (Figure 11 of the Supporting Information). The ω torsional angle of the Neu5Ac-α(2→6)-Gal-β disaccharide has previously been discussed by Sabesan et al.[40] for similar di- or trisaccharides. They propose that two states are allowed for ω torsional angles, located approximately at −50° and 180°, corresponding to antiperiplanar (gt conformer) and synclinal (tg) for H5−H6(R) (Gal1), respectively. These states have characteristic $J_{H5-H6(R)}$ couplings, ∼10 Hz for the former and a smaller value of ∼3.5 Hz for the latter. We observed a coupling constant of 10 Hz for $J_{H5-H6(R)}$ in LSTc, which according to Sabesan et al.[40] corresponds to the antiperiplanar H5−H6(R) Gal1 conformer, which supports the values determined by the MD simulation.

As we move along the glycans, from the nonreducing end to the reducing end, the glycosidic linkages exhibit progressively decreasing disparities between the glycosidic torsional state positions and populations (Figure 3). The ϕ_2 and ψ_2 distribution for LSTc and LSTa (Figure 3 and Table 2) qualitatively agree with those found previously by Sabesan et al.[40] for disaccharides Gal-β(1→3)-GlcNAc and Gal-β(1→4)-GlcNAc, which are model compounds for LSTa and LSTc, respectively.

The torsional states determined for LSTa broadly agree with those described by Xu et al.,[17] while the torsional angles determined for LSTc are less consistent with the data of the aforementioned work. Xu et al.[17] found the following additional LSTc torsional states: ∼180° and ∼180° for ϕ_1 and ψ_1, ∼60° and ∼180° for ϕ_2 and ψ_2, ∼30° and ∼160° for ϕ_3 and ψ_3, and ∼50° and ∼180° for ϕ_4 and ψ_4, respectively. We predict a state in the first glycosidic linkage of LSTc located at 56° and 174°, which was not observed by Xu et al.,[17] but this state is seen in the conformational map obtained at GlycoMapsDB for Neu5Ac-α(2→6)-Gal.[41] This discrepancy is possibly related to the use of a different starting conformation for LSTc during

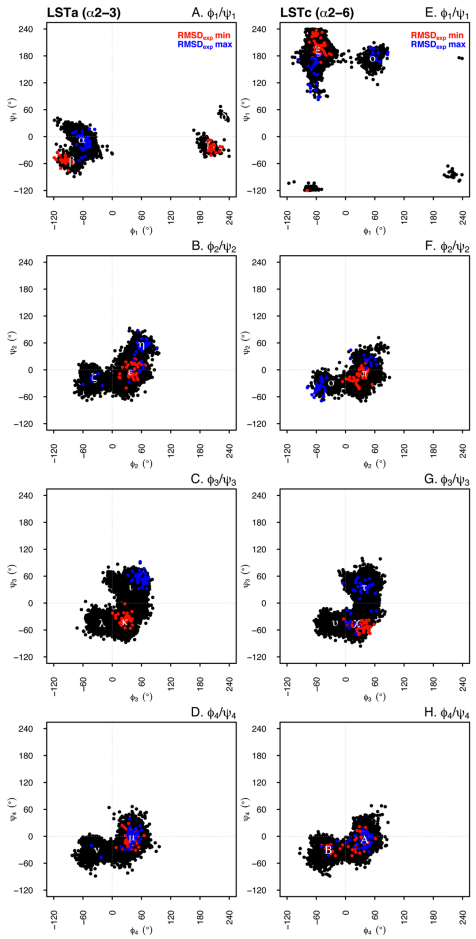


Figure 3. Torsional states visited during the 100 ns simulation time for LSTa (left) and LSTc (right). For each glycan, the dihedral angle pairs [ϕ_1 and ψ_1 (A and E), ϕ_2 and ψ_2 (B and F), ϕ_3 and ψ_3 (C and G), and ϕ_4 and ψ_4 (D and H)] are shown. The red dots represent conformers that better fit the experimental NMR NOESY data (rmsd$_{exp}$ min set of conformations), while blue dots correspond to the conformers that exhibit poor agreement with the experimental constraints (rmsd$_{exp}$ max). The positions and populations of the torsional states are listed in Table 2.

the MD simulation, which prevents an ensemble of conformations from being visited as a consequence of the potential energy walls restricting exploration of the conformational space. To partially address this limitation, shorter MD

Table 2. Torsional States of LSTa and LSTc as Determined by Molecular Dynamics

linkage		ϕ (deg)	ψ (deg)	no.	%
		LSTa (α2→3) Torsional angles			
1	α	−62	−8	8982	89.8
	β	−83	−55	770	7.7
	γ	196	−19	227	2.3
	δ	230	47	21	0.2
2	ε	37	−5	6908	69.1
	ζ	−37	−18	2402	24.0
	η	60	54	690	6.9
3	ι	38	35	5534	55.3
	κ	26	−43	3134	31.3
	λ	−21	−44	1332	13.3
4	μ	40	−6	7732	77.3
	ν	−32	−31	2268	22.7
		LSTc (α2→6) Torsional angles			
1	ξ	−57	190	9622	96.2
	o	56	174	378	3.8
2	π	40	−9	9024	90.2
	σ	−30	−30	976	9.8
3	τ	39	34	4108	41.1
	χ	23	−46	4118	41.2
	υ	−21	−45	1774	17.7
4	A	39	−7	5840	58.4
	B	−36	−30	4160	41.6

simulations (20 ns) were performed at higher temperatures (395 and 495 K) using the same starting conformation at each temperature (Figures 12 and 13 of the Supporting Information). At these higher temperatures, the missing LSTc torsional states in linkages ϕ_1 and ψ_1 and ϕ_4 and ψ_4 started to be populated (Figure 12 of the Supporting Information). Interestingly, for both LSTc and LSTa, the first glycosidic torsional state (ϕ_1 and ψ_1) reproduces qualitatively in position and shape the torsional angle map of the Neu5Ac-α(2→6/3)-Gal glycosidic junction shown in GlycoMapsDB.[41]

A parameter used to monitor the evolution of a MD simulation is the root-mean-square displacement (rmsd) of the atoms in a molecule (determined using the conformation at time zero as a reference), which illustrates the different conformational states that the molecule assumes during the simulation, as compared to the starting state. Unsurprisingly, the conformational states assumed by LSTc and LSTa are related to the torsional angle states of their glycosidic linkages. During the simulation, LSTc moves between two conformational states, rmsd states located at ∼2.10 and ∼3.50 Å (Figure 14 of the Supporting Information), while LSTa assumes a single broad distribution of related conformational states (Figure 15 of the Supporting Information) centered at ∼2.8 Å. The conformers that populate the different rmsd states observed for LSTc, located at rmsd values of ∼2.10 and ∼3.50 Å, have stark differences in the third glycosidic linkage (ϕ_3 and ψ_3), GlcNAc-β(1→3)-Gal-β (Figure 16 of the Supporting Information). This correlation is also true for LSTa; a subset of conformers located at rmsd values of 2.65 and 3.80 Å (±10 Å) also exhibit significant differences in the third glycosidic linkage (Figure 17 of the Supporting Information). This result is interesting for the case of LSTc, as it suggests that the third glycosidic linkage is the most conformationally mobile junction and not the first, extended linkage Neu5Ac-α(2→6)-Gal, as might be assumed. This could be due to the first residue forming hydrogen bonds with the rest of the molecule, the result being that the first linkage is static compared to the third.

NMR Data and MD Simulation Comparison. To compare the conformers derived by MD simulation and the experimental NMR data, an $rmsd_{exp}$ was derived from five experimentally estimated inter-residue proton−proton distances (see eq 2 and Table 1) and the corresponding MD predicted distances for each structure in the MD trajectories (see Experimental Procedures). The smaller the $rmsd_{exp}$ value, the more closely the conformer matches the NMR data.

The MD-generated conformers of LSTc have two distinct $rmsd_{exp}$ states. The first state located at 0.64 Å is very narrow, and these conformers match the NMR data well; this population ends at 1.05 Å, and thereafter, there is a very broad declining population of conformers with increasing disparity between the NMR-estimated inter-residue proton−proton distances and MD-predicted distances (Figure 18 of the Supporting Information). The conformers that populate these states have distinct glycosidic torsional states, with the conformers that more closely match the NMR data, the population with the smaller $rmsd_{exp}$ values (<1.05 Å), belonging to the following glycosidic torsional angle states: −57° and 190° for ϕ_1 and ψ_1 and 39° and −9° for ϕ_2 and ψ_2, respectively (clusters ξ and π, respectively, Figure 3E,F and Figure 19 of the Supporting Information). LSTa also has two $rmsd_{exp}$ populations, located at 0.75 and 0.97 Å (Figure 20 of the Supporting Information); again the conformers that comprise these different populations belong to differing torsional angle states. In this circumstance, the differentiating linkage is closer to the reducing end of the glycan, ϕ_3 and ψ_3; the MD-generated conformers that are in closer agreement to the NMR data are members of the torsional angle states located at −21° and −45° and 23° and −46° for ϕ_3 and ψ_3, respectively (clusters λ and κ, respectively, Figure 3C and Figure 21 of the Supporting Information). It is interesting to note that on the whole the MD conformers of LSTa more closely match the NMR data, with the whole population falling within the $rmsd_{exp}$ range of 0−1.5 Å, whereas the LSTc conformers have $rmsd_{exp}$ values spanning the range of 0−5 Å.

To further compare the experimental NMR data with the MD-determined structures, two subsets of conformers were chosen for each glycan, a subset containing the conformers with the 40 smallest $rmsd_{exp}$ values ($rmsd_{exp}$ min) (0.24−0.33 Å for LSTc and 0.12−0.43 Å for LSTa) and a subset comprising the structures with the 40 largest $rmsd_{exp}$ values ($rmsd_{exp}$ max) (1.26−1.41 Å for LSTa and 3.83−4.40 Å for LSTc). The 2D NOESY spectra were then simulated for these selected structures (see Experimental Procedures), with the calculated intra- and inter-residue proton−proton NOE signals then being compared (Table 4 of the Supporting Information). The H1−H3, H1−H5, and H3−H5 intraresidue NOEs signals are qualitatively reproduced for LSTc residues Neu5Ac, Gal1, GlcNAc, and Gal2, while for LSTa, the signals are reproduced for only Gal1, GlcNAc, and Gal2. Interestingly, the agreement between the simulated and experimental intraresidue 2D NOE signals is similar for both $rmsd_{exp}$ min and $rmsd_{exp}$ max conformer subsets, because the $rmsd_{exp}$ values were determined using inter-residue distances and not intraresidue distances. In fact, the H1 Gal1−H4 GlcNAc, H1 GlcNAc−H3 Gal2, H1 Gal2−H4 Glc, and CH_3 GlcNAc−H5 Neu5Ac inter-residue distances compare well for the LSTc $rmsd_{exp}$ min conformer subset, while the $rmsd_{exp}$ max conformers reproduce only the H1 Gal2−H4 Glc distance. This trend is also true for LSTa,

with the $rmsd_{exp}$ min set of structures reproducing the H3 Gal1−$H3_{ax}$ Neu5Ac, H1 Gal1−H3 GlcNAc, H1 GlcNAc−H3 Gal2, and H1 Gal2−H4 Glc inter-residue distances well and the $rmsd_{exp}$ max conformers replicating adequately only the experimentally determined H1 Gal2−H4 Glc distance (Table 4 of the Supporting Information). This illustrates the ability of the $rmsd_{exp}$ minimization procedure to select conformers from the MD simulation trajectory, providing a means of determining that quality of the MD simulation and allowing the structures that more closely match the experimental data to be extracted. To make a broader comparison of the modeled and measured data, with MD conformers not being chosen by experimental constraints, the last 60 geometries of the MD simulation for LSTa and LSTc were used to generate NOE magnitudes (Table 4 of the Supporting Information). For both glycans, the agreement between the NOE magnitudes derived from MD and experimental data is qualitatively good, but the NOE magnitudes for the $rmsd_{exp}$ min MD conformer subset more closely match the experimental data than the NOE magnitudes derived from the MD structures, which were chosen without experimental constraints. It appears that the first glycosidic linkage, Neu5Ac-α(2→6/3)-Gal, for both glycans is the least well described by the MD simulation, while the remaining interglycosidic NOEs are better modeled in LSTa. This could be due to the isotropic motion approximation introduced by NOEPROM to generate the theoretical NOE magnitudes, with LSTc exhibiting stronger motional anisotropy than LSTa, which is observed in the NMR relaxation measurements.

The torsional angles for the $rmsd_{exp}$ min and $rmsd_{exp}$ max conformer subsets are shown in Figure 3 for LSTa and LSTc, respectively (Table 5 of the Supporting Information). $rmsd_{exp}$ min structures are colored red, while the $rmsd_{exp}$ max conformers are colored blue. Both of the conformer subsets for LSTc and LSTa show significant differences in their torsional angle states; the ϕ_4 and ψ_4 linkage for both glycans is the only linkage where there is little divergence. Interestingly, the LSTa $rmsd_{exp}$ min conformers have two ϕ_1 and ψ_1 states, the linkage between Neu5Ac and Gal1, which are located at −100° and −53° and 207° and −24°, corresponding to clusters β and γ, respectively. The conformers in the state centered at −100° and −53° are in better agreement with the NOE values for the $H3_{ax}$ Neu5Ac−H3 Gal1 distance (Table 4 of the Supporting Information). Sabesan et al. also observed a similar NOE, between $H3_{ax}$ Neu5Ac and H3 Gal1, for the di- and trisaccharide structures.[40] The glycosidic torsional angles of the conformers that more closely fit the NMR data are as follows: −55° and 197° for ϕ_1 and ψ_1, 28° and −15° for ϕ_2 and ψ_2, 34° and 54° for ϕ_3 and ψ_3, and 11° and 15° for ϕ_4 and ψ_4, respectively, for LSTc and 207° and −24° for ϕ_1 and ψ_1, 41° and −1° for ϕ_2 and ψ_2, 27° and 34° for ϕ_3 and ψ_3, and 38° and −5° for ϕ_4 and ψ_4, respectively, for LSTa. To corroborate this result, selected long-range coupling constants, $^3J_{CH}$, across the glycosidic linkages ($J\phi$ for H1−C1−O*x*−C*x* and $J\psi$ for C1−O*x*−C*x*−H*x*) were measured; this allowed an indirect estimation of the glycosidic torsional angles for both glycans. The measured coupling constants (in hertz) starting from the first torsion (ϕ_1 and ψ_1) are as follows: $J\psi_1$ = 4.4, $J\phi_2$ = 2.6, $J\psi_2$ = 5.4, $J\phi_3$ = 4.2, $J\psi_3$ = 4.5, and $J\phi_4$ = 4.1 for LSTa and $J\psi_1$ = 2.6, $J\phi_2$ = 3.6, $J\phi_3$ = 4.1, $J\psi_3$ = 4.4, and $J\phi_4$ = 3.4 for LSTc, respectively. Calculated pairs of couplings ($J\phi$ and $J\psi$), determined using the $rmsd_{exp}$ min subset of conformers, in accordance to the Karplus-type equation,[42] are reported in Table 7 of the Supporting Information. The only measured long-range coupling that does not agree with the $rmsd_{exp}$ min selected subset of conformers is the ϕ_1 and ψ_1 coupling for LSTa; the measured coupling ($J^{exp}\psi_1$ = 4.4 Hz) is in better agreement with the glycosidic torsional angles ϕ_1 and ψ_1 of −99° and −53°, respectively, and not ϕ_1 and ψ_1 values of 207° and −24°, respectively, which were determined by the NMR-selected MD conformers. J-HMBC spectra of LSTa and LSTc, overlaid with HSQC spectra, can be found in Figure 6 of the Supporting Information.

Analyses of the Shape and Topology of LSTc and LSTa. The global shape of the two oligosaccharides is described by the topological θ angle, as defined by Xu et al.[17] (see Experimental Procedures). Chandrasekaran et al. previously proposed that the θ angle of LSTc and LSTa defines their conformation, with LSTa adopting a linear form, coined a cone, which is defined as having a θ angle >110°, while LSTc, with its extra degree of freedom at the nonreducing end, Neu5Ac-α(2→6)-Gal-β(1→4)-, forms a bent conformation, termed the umbrella form, defined as having a θ angle of <110°.[16]

The LSTc conformers determined by MD have two θ angle populations; the much larger population is located at 87°, and the second smaller population starts above 109° (Figure 9 of the Supporting Information). These differences in θ values also correspond to differences in the torsional angle states for the linkage between Gal1 and GlcNAc; the subset of conformers with a θ angle centered at 87° correlates with the torsional angle state with values for ϕ_2 and ψ_2 of 39° and −9°, respectively, while the structures with a θ angle of >88° populate torsional states with values for ϕ_2 and ψ_2 of 39° and −9° and −30° and −30°, respectively (Figure 22 of the Supporting Information). This indicates that the second glycosidic linkage is important in the "opening and closing" of the umbrella conformation proposed for LSTc and not the first glycosidic linkage, which might be assumed because of the α2→6 linkage between Neu5Ac and Gal1.

LSTa also has two θ angle populations, with the more populated state centered at 153° and the less populated state located at 119° (Figure 10 of the Supporting Information). As for LSTc, the LSTa conformers that belong to the different θ angle populations correlate strongly with the torsional states of the second glycosidic linkage, ϕ_2 and ψ_2 (Figure 23 of the Supporting Information). The θ angle population centered at 153° correlates with the torsional state with values for ϕ_2 and ψ_2 of 37° and −5°, respectively, while the population located at 119° correlates with the torsional state with values for ϕ_2 and ψ_2 of −37° and 18°, respectively. These results highlight that the conformational and subsequent dynamic properties of both LSTc and LSTa are linked strongly to the second glycosidic linkage.

The LSTc conformers that more closely fit the NMR data ($rmsd_{exp}$ min structures) have an average θ angle of 87° (standard deviation of 4°), while the corresponding LSTa conformers have an average θ angle of 159° (standard deviation of 5°). These values, for the NMR-selected MD-generated conformers, agree with the results published by Xu et al.[17] and the hypothesis proposed by Chandrasekaran et al.[16]

Principal Component Analysis of the LSTa and LSTc MD Trajectories. Principal component analysis (PCA) has been applied on the MD-simulated trajectory to extract the independent mode of motion of both oligosaccharides, allowing an easier correlation with its structural properties. PCA has been a powerful tool for the analysis of protein[35,43] and

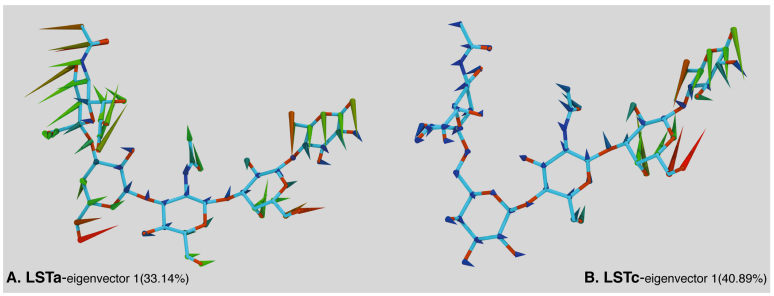


Figure 4. Porcupine plot representations of the first eigenvectors for LSTa (A) and LSTc (B). The direction, size, and color of the spines are proportional to the direction and magnitude of the motion of the atoms. The illustration shows LSTa moving in a bowlike motion while LSTc is moving in two independent sections: NeuSAc synchronized with the GlcNAc and the reducing-end disaccharide as a single unit. In LSTc, Gal1 is the least mobile residue.

glycan[44] molecular dynamics. The convergence of the distinct ensembles of motion was observed by determining the rmsip of the MD trajectory eigenvectors, comparing the eigenvectors generated by PCA of the whole MD tracjectory (100 ns) to the PCA of the trajectory at different time points (increasing steps of 1 ns). The rmsip provides a measure of the similarity of the "essential modes" extracted; if the value is zero, then they are dissimilar, and if the value is one, the subspaces have converged. The rmsip for both LSTc and LSTa is ≥0.9 for at least 50% of the trajectory; Amadei et al.[35] opined that values of >0.7 provide evidence of good convergence (Figure 24 of the Supporting Information).

Here, only the first four components generated by PCA for LSTa and LSTc are considered. They explain 73.8 and 66.2% of the total variance, respectively. The PCA scree, loading, and normalized eigenvector plots can be found in Figures 25−31 of the Supporting Information. The normalized eigenvector plots indicate the contribution each atom makes to the extracted components, i.e., whether an atom, or a group of atoms, is involved in the independent dynamic motions isolated by PCA.

The independent modes of motion described by component one for LSTc have the nonreducing end of the glycan close to stationary compared to the other residues, while the reducing end (Gal2-Glc) rotates around their glycoside linkages, moving as a single unit. This is in agreement with the rmsd analysis of the LSTc MD trajectory. Figure 4 provides an illustration of this mode of motion, "porcupine" representations of the first eigenvector for LSTc and LSTa [supplementary data contain graphical representations of all four components (Figures 30 and 31 of the Supporting Information)]. The three other distinct modes of motion extracted for LSTc show both the nonreducing and reducing ends of the molecule are in motion, with the reducing end disaccharide (Gal2-Glc) moving as one unit and the sialyl group of NeuSAc moving toward the *N*-acetyl group of GlcNAc. The least mobile residue is Gal1, which exhibits the smallest variation of the four components examined, which is illustrated in the normalized eigvector figure for LSTc (Figure 28 of the Supporting Information). These results are consistent with a small θ angle and a hydrogen bonding network between GlcNAc and NeuSAc of LSTc, while the fact that the reducing end of the molecule moves in unison is indicative of a cellobiose-like hydrogen bonding network between Glc and Gal2. This dynamic description agrees with the results of the T_1 and T_2 relaxation measurements.

The independent modes of motion extracted for LSTc correlate with structural parameters that have been previously discussed; the motion described by component 1 correlates with the rmsd of the conformers and is therefore related to the different torsional states in the third glycosidic linkage, GlcNAc-β(1→3)-Gal-β (Figures 16 and 32 of the Supporting Information). The molecular motion illustrated by the second component correlates with the structures that belong to the different θ angle populations, which are strongly tied to differences in the second glycosidic linkage (Gal-β(1→4)-GlcNAc) (Figures 22 and 34 of the Supporting Information). Finally, the dynamic that distinguishes whether a conformer agrees well with the experimental NMR data, $rmsd_{exp}$, correlates with component 3, which is related to the different torsional states at glycosidic linkages 1 and 2 (Figures 19 and 36 of the Supporting Information).

The distinct modes of motion isolated for LSTa are very different from those of LSTc. LSTa has "bowlike" dynamics for the four components examined, the whole molecule moving in unison; this is particularly true of the motion isolated for component 1 (Figure 29 of the Supporting Information). Again, the modes of motion isolated by PCA correlate with previously discussed structural parameters. A mixture of components distinguishes the conformers with different rmsd values, for example, the mixture of components 1 and 2 and also components 1 and 4. These dynamics, which are described by the mixtures of components, are generated by specific states in the second and third glycosidic linkages (Figures 17 and 33 of the Supporting Information). Component 1 correlates strongly with the θ angle of the conformers, which in turn is related to the torsional state of ϕ_2 and ψ_2 (Figures 23 and 35 of the Supporting Information). Component 2 is the independent motion that discriminates the structures that fit well with the experimental NMR data from those that do not; high or low $rmsd_{exp}$ values are associated with differences in the torsional

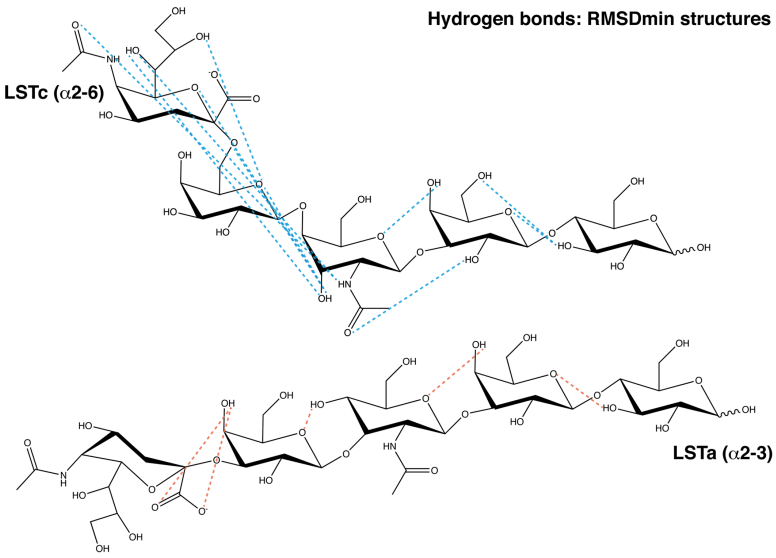


Figure 5. Proposed hydrogen bond network for LSTc and LSTa $rmsd_{exp}$ min conformers. The hydrogen bond network at the nonreducing end of each glycan is strongly affected by the different linkage between Neu5Ac and Gal1 [$\alpha(2\rightarrow3)$ for LSTa and $\alpha(2\rightarrow6)$ for LSTc]. The acceptor and donor atoms lie within 2.6 Å of each other, and the X−H···Y (X is the donor and Y the acceptor) angle was greater than 110°, tending toward 180°. Predicted hydrogen bonds for both LSTa and LSTc appear in Table 6 of the Supporting Information.

states of ϕ_3 and ψ_3 (Figures 20 and 37 of the Supporting Information).

Inter-Residue Hydrogen Bond Analysis. To further understand the consequences and driving forces behind the different conformations assumed by LSTc and LSTa, the possible inter-residue hydrogen bonds were determined for the $rmsd_{exp}$ min and $rmsd_{exp}$ max conformer subsets. The analysis employed distance and angular restrictions; the distance between the donated hydrogen and the acceptor was ≤2.6 Å, and the X−H···Y (X is the donor and Y the acceptor) angle was greater than 110°, tending toward 180°. Here we will discuss the proposed hydrogen bonds for the $rmsd_{exp}$ min subset of conformers, the structures that more accurately represent the experimental data; the hydrogen bond analysis for the $rmsd_{exp}$ max conformers, for comparison, is shown in Table 6 and Figure 38 of the Supporting Information.

For both glycans, several possible hydrogen bonds were determined for the $rmsd_{exp}$ min conformers, 5 for LSTa and 11 for LSTc, illustrated in Figure 5, while the populations of each are reported in Table 6 of the Supporting Information. Both LSTc and LSTa have potential hydrogen bonds that offer an explanation for the tandem motion of the reducing end residues Glc and Gal2 isolated by PCA: H3O Glc−O5 Gal2 (percentage of conformers, 62.5% for LSTc and 80% for LSTa) (cellobiose-like), H4O Gal2−O5 GlcNAc (82.5% for LSTc and 22.5% for LSTa), H2O Gal2−O2N GlcNAc (42.5% for LSTc), and H3O Glc−O6 Gal2 (12.5% for LSTc). The cellobiose-like hydrogen bond between H3O of Glc and O5 of Gal2 has previously been observed in a structurally related molecule, Neu5Ac-$\alpha(2\rightarrow6)$-Gal-$\beta(1\rightarrow4)$-Glc, sialyl-$\alpha(2\rightarrow6)$-lactose (S6L).[45]

The hydrogen bond networks present at the nonreducing end of either glycan are a direct consequence of their θ angles. The nonreducing end hydrogen bond network for LSTc, which has a bent form in solution (θ = 86° for the selected conformers), originates from GlcNAc, with two predicted bonds to Gal1 [H3O GlcNAc−O5 Gal1 (100%) and H3O GlcNAc−O6 Gal1 (25.0%)] and five others spanning Gal1 to Neu5Ac [O3 GlcNAc−H7O Neu5Ac (55.0%), O5 Gal1−H8O Neu5Ac (20.0%), H2N GlcNAc−O5N Neu5Ac (15.0%), H3O GlcNAc−O6 Neu5Ac (12.5%), and H2N GlcNAc−O7 Neu (12.5%)] (Figure 5 and Table 6 of the Supporting Information). A hydrogen bond similar to the O3 GlcNAc−H7O Neu5Ac hydrogen bond is observed in S6L, between O3 of Glc and H7O of Neu5Ac.[45] Because of the bowlike backbone conformation of LSTa, its hydrogen bonds can be formed only between directly adjacent residues: H4O GlcNAc−O5 Gal1 (92.5%), H4O Gal1−O1A/O1B (acid group) Neu5Ac (52.5%), H4O Gal2−GlcNAc (22.5%), and H3O Glc−O5 Gal2 (80.0%) (Figure 5 and Table 6 of the Supporting Information). All of the proposed hydrogen bonds are listed in

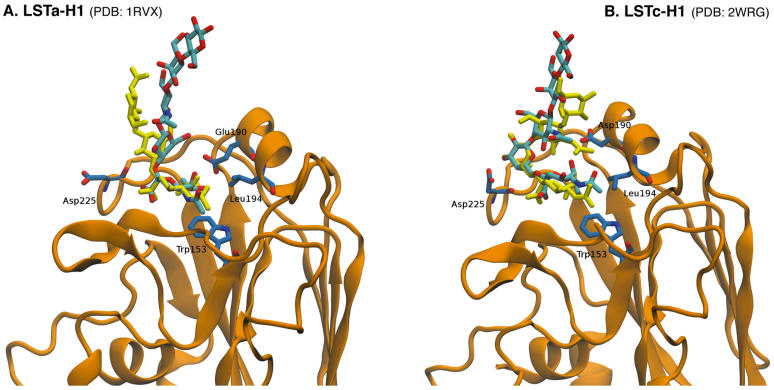


Figure 6. Comparison of the MD -generated LSTa (A) and LSTc (B) conformers (structures with the smallest $rmsd_{exp}$ value, $rmsd_{exp}$ min) and crystal structures of LSTa and LSTc bound to hemagglutinin from H1. The conformers generated via MD simulations are colored cyan (C atoms) and red (O atoms), while the conformers determined by X-ray crystallography are colored yellow. The rmsd values between the MD simulation-generated structures and the cocrystallized glycans are 6.2 and 6.5 Å for LSTa and LSTc, respectively.

Table 6 of the Supporting Information, while Figure 5 is a schematic showing the major proposed inter-residue hydrogen bonds.

■ DISCUSSION

This study has identified the distinct conformational, topological, and dynamic properties of the glycan receptors [two pentasaccharides, LSTc (Neu5Ac-α(2→6)-Gal-β(1→4)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc) and LSTa (Neu5Ac-α(2→3)-Gal-β(1→3)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc] of influenza A viruses in solution, i.e., without its receptor. The study involved utilizing extensive NMR measurements, including the complete ^{1}H and ^{13}C assignment for each glycan, MD simulations, and the numerical analysis of the data generated, to provide not only information concerning the conformation in solution but also, for the first time, details of its overall dynamic properties that could be related to the receptor binding ability.

Conformers from the MD simulations were selected using a $rmsd_{exp}$ function, comparing NOE enhancement-derived inter-residue proton−proton distances and the equivalent predicted distances. Using this approach, two subsets of conformers were selected, in good ($rmsd_{exp}$ min) or poor ($rmsd_{exp}$ max) agreement with the experimental data. In particular, the conformers that were in better agreement with the experimental restraints, $rmsd_{exp}$ min, generated 2D NOESY spectra, using the multispin approximation (NOEPROM), in good agreement with the values measured at 600 and 900 MHz. These results were subsequently corroborated by experimental $^{n}J_{C-H}$ heteronuclear coupling constants, which were measured across the glycosidic linkages for LSTc and LSTa; these results supported the concept of using NMR experimental constraints to select conformers from MD simulation trajectories. LSTc exhibits a bent shape characterized by a topological angle θ of 87°, whereas LSTa has a more linear form and therefore a greater θ angle of 159°, a structural difference in agreement with the previous work of Chandrasekaran[16] and Xu et al.[17] that concurs with the glycan ligand cocrystallized with HA.[15]

As a consequence of these structural differences, the nonreducing end terminal Neu5Ac of LSTc could form an intricate hydrogen bond network with GlcNAc, which is not the case for its isomer, LSTa, because hydrogen bond analysis revealed interactions with only directly adjacent residues. PCA of the glycan trajectories extracted the independent modes of motion that differentiate the structures, with LSTa moving like a flexing bow and the reducing and nonreducing disaccharides moving around the GlcNAc fulcrum, whereas the dynamics of LSTc are different from those of LSTa, the sialyl group of Neu5Ac moving in tandem with the *N*-acetyl group of GlcNAc, which moves in a manner that is independent of the reducing end disaccharide.

In Figure 6, the structures of LSTc and LSTa extracted from the respective $rmsd_{exp}$ min sets are superimposed on the Neu5Ac and Gal1 residues in the cocrystallized glycan−HA complexes [PDB entries 2WRG[16] (panel B) and 1RVX[2] (panel A)]. The rmsd values between the superimposed structures and the cocrystallized ligand, considering the three nonreducing end residues (i.e., Neu5Ac, Gal1, and GlcNAc), are 6.5 and 6.2 Å for LSTa and LSTc, respectively. These results illustrate how the NMR-selected conformation for LSTc and LSTa in solution, i.e., determined without HA, matches qualitatively the cocrystallized ligand bound to HA and also suggests that the respective sugar conformation does not significantly change upon interaction with HA. This observation supports the hypothesis that HA recognizes the solution conformation of the glycan, instead of the ligand and receptor undergoing a "bind and fit" process. This study contributes to an improved

understanding of the structural basis for the glycan receptor specificity of HA in influenza.

■ ASSOCIATED CONTENT

Supporting Information

Supporting NMR data and analyses of the MD simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Istituto di Ricerche Chimiche e Biochimiche "G. Ronzoni", Via Giuseppe Colombo, 81, Milano, 20133 Italy. E-mail: guerrini@ronzoni.it. Telephone: +39-2-70641627.

Present Addresses

∥G.L.S.: Departamento de Bioquimica e Biologia Molecular, Universidade Federal do Paraná, CP 19.046, CEP81.531-980 Curitiba, PR, Brazil.

⊥T.R.R.: Diamond Light Source Ltd., Diamond House, Harwell Science & Innovation Campus, Didcot, Oxfordshire OX11 0DE, U.K.

Author Contributions

G.L.S., S.E., and T.R.R. made complementary and equal contributions to this work.

Funding

We thank Finlombardia SPA 'Fondo per la promozione di Accordi Istituzionali' and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for their financial support. This work was also funded in part by the National Institutes of Health (R37 GM057073-13) and the National Research Foundation supported Interdisciplinary Research group in Infectious Diseases of SMART (Singapore MIT Alliance for Research and Technology). The 900 MHz spectra were recorded at the SONNMR Large Scale Facility in Utrecht, by the financial support of the Access to Research Infrastructures activity in the seventh Framework Programme of the EC (contract no. 261863, EU-NMR).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are very grateful for access to the NMR spectrometer (900 MHz) at the Utrecht NMR Facility and the assistance of Dr. Hans Wienk.

■ ABBREVIATIONS

HA, hemagglutinin; LSTa, Neu5Ac-α(2→3)-Gal-β(1→3)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc; LSTc, Neu5Ac-α(2→6)-Gal-β(1→4)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc; MD, molecular dynamics; Gal, galactose; Neu5Ac, *N*-acetylneuraminic acid; GlcNAc, *N*-acetylglucosamine; Glc, glucose; rmsip, root-mean-square internal product; rmsd, root-mean-square displacement; $rmsd_{exp}$, comparison of root-mean-square displacements between experimental and theoretical data; PCA, principal component analysis.

■ REFERENCES

(1) Eisen, M. B., Sabesan, S., Skehel, J. J., and Wiley, D. C. (1997) Binding of the influenza A virus to cell-surface receptors: Structures of five hemagglutinin-sialyloligosaccharide complexes determined by X-ray crystallography. *Virology* 232, 19−31.

(2) Gamblin, S. J., Haire, L. F., Russell, R. J., Stevens, D. J., Xiao, B., Ha, Y., Vasisht, N., Steinhauer, D. A., Daniels, R. S., Elliot, A., Wiley, D. C., and Skehel, J. J. (2004) The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science 303*, 1838−1842.

(3) Ha, Y., Stevens, D. J., Skehel, J. J., and Wiley, D. C. (2001) X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. *Proc. Natl. Acad. Sci. U.S.A. 98*, 11181−11186.

(4) Stevens, J., Blixt, O., Tumpey, T. M., Taubenberger, J. K., Paulson, J. C., and Wilson, I. A. (2006) Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science 312*, 404−410.

(5) Weis, W., Brown, J. H., Cusack, S., Paulson, J. C., Skehel, J. J., and Wiley, D. C. (1988) Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. *Nature* 333, 426−431.

(6) Ibricevic, A., Pekosz, A., Walter, M. J., Newby, C., Battaile, J. T., Brown, E. G., Holtzman, M. J., and Brody, S. L. (2006) Influenza virus receptor specificity and cell tropism in mouse and human airway epithelial cells. *J. Virol. 80*, 7469−7480.

(7) Shinya, K., Ebina, M., Yamada, S., Ono, M., Kasai, N., and Kawaoka, Y. (2006) Avian flu: Influenza virus receptors in the human airway. *Nature 440*, 435−436.

(8) van Riel, D., Munster, V. J., de Wit, E., Rimmelzwaan, G. F., Fouchier, R. A., Osterhaus, A. D., and Kuiken, T. (2007) Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals. *Am. J. Pathol. 171*, 1215−1223.

(9) Jayaraman, A., Pappas, C., Raman, R., Belser, J. A., Viswanathan, K., Shriver, Z., Tumpey, T. M., and Sasisekharan, R. (2011) A single base-pair change in 2009 H1N1 hemagglutinin increases human receptor affinity and leads to efficient airborne viral transmission in ferrets. *PLoS One 6*, e17616.

(10) Kumari, K., Gulati, S., Smith, D. F., Gulati, U., Cummings, R. D., and Air, G. M. (2007) Receptor binding specificity of recent human H3N2 influenza viruses. *Virol. J. 4*, 42.

(11) Matrosovich, M. N., Matrosovich, T. Y., Gray, T., Roberts, N. A., and Klenk, H. D. (2004) Human and avian influenza viruses target different cell types in cultures of human airway epithelium. *Proc. Natl. Acad. Sci. U.S.A. 101*, 4620−4624.

(12) Rogers, G. N., Paulson, J. C., Daniels, R. S., Skehel, J. J., Wilson, I. A., and Wiley, D. C. (1983) Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature 304*, 76−78.

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

(14) Shriver, Z., Raman, R., Viswanathan, K., and Sasisekharan, R. (2009) Context-specific target definition in influenza A virus hemagglutinin-glycan receptor interactions. *Chem. Biol. 16*, 803−814.

(15) Bewley, C. A. (2008) Illuminating the switch in influenza viruses. *Nat. Biotechnol. 26*, 60−62.

(16) Chandrasekaran, A., Srinivasan, A., Raman, R., Viswanathan, K., Raguram, S., Tumpey, T. M., Sasisekharan, V., and Sasisekharan, R. (2008) Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nat. Biotechnol. 26*, 107−113.

(17) Xu, D., Newhouse, E. I., Amaro, R. E., Pao, H. C., Cheng, L. S., Markwick, P. R., McCammon, J. A., Li, W. W., and Arzberger, P. W. (2009) Distinct glycan topology for avian and human sialopentasaccharide receptor analogues upon binding different hemagglutinins: A molecular dynamics perspective. *J. Mol. Biol. 387*, 465−491.

(18) R Development Core Team (2012) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna.

(19) Lewis, I. A., Schommer, S. C., and Markley, J. L. (2009) rNMR: Open source software for identifying and quantifying metabolites in NMR spectra. *Magn. Reson. Chem. 47* (Suppl. 1), S123−S126.

(20) Sarkar, D. (2008) *Lattice: Multivariate Data Visualization with R*, Springer, New York.

(21) Neuhaus, D., and Williamson, M. (1989) *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VHC Publishers Inc., New York.
(22) Angulo, J., Hricovini, M., Gairi, M., Guerrini, M., de Paz, J. L., Ojeda, R., Martin-Lomas, M., and Nieto, P. M. (2005) Dynamic properties of biologically active synthetic heparin-like hexasaccharides. *Glycobiology 15*, 1008−1015.
(23) Hricovini, M., and Torri, G. (1995) Dynamics in aqueous solutions of the pentasaccharide corresponding to the binding site of heparin for antithrombin III studied by NMR relaxation measurements. *Carbohydr. Res. 268*, 159−175.
(24) Poveda, A., Martin-Pastor, M., Bernabe, M., Leal, J. A., and Jimenez-Barbero, J. (1998) Solution conformation and dynamics of a fungal cell wall polysaccharide isolated from *Microsporum gypseum*. *Glycoconjugate J. 15*, 309−321.
(25) Woods Group (2005−2012) GLYCAM *Web*.
(26) Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem. 29*, 622−655.
(27) Jakalian, A., Jack, D. B., and Bayly, C. I. (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem. 23*, 1623−1641.
(28) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926.
(29) Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B. P., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossvai, I., Wong, K. F., Paesani, F., Vanicek, J., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, J., Wang, J., Hsieh, M.-J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2010) *AMBER 11*, University of California, San Francisco.
(30) Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem. 26*, 1781−1802.
(31) Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A., and Caves, L. S. (2006) Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics 22*, 2695−2696.
(32) Martin-Pastor, M. (2005) *NOEPROM*.
(33) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graphics 14*, 27−38.
(34) Schulz, R. (2009) http://www.ks.uiuc.edu/Research/vmd/mailing_list/vmd-l/att-15365/PorcupinePlot.tcl.
(35) Amadei, A., Linssen, A. B., and Berendsen, H. J. (1993) Essential dynamics of proteins. *Proteins 17*, 412−425.
(36) Turner, C. J., Connolly, P. J., and Stern, A. S. (1999) Artifacts in sensitivity-enhanced HSQC. *J. Magn. Reson. 137*, 281−284.
(37) Poppe, L., and Van Halbeek, H. (1992) The rigidity of sucrose: Just an illusion? *J. Am. Chem. Soc. 114*, 1092−1094.
(38) Poppe, L., van Halbeek, H., Acquotti, D., and Sonnino, S. (1994) Carbohydrate dynamics at a micellar surface: GD1a headgroup transformations revealed by NMR spectroscopy. *Biophys. J. 66*, 1642−1652.
(39) Raab, M., and Tvaroska, I. (2011) The binding properties of the H5N1 influenza virus neuraminidase as inferred from molecular modeling. *J. Mol. Model. 17*, 1445−1456.
(40) Sabesan, S., Bock, K., and Paulson, J. C. (1991) Conformational analysis of sialyloligosaccharides. *Carbohydr. Res. 218*, 27−54.
(41) Frank, M., Lutteke, T., and von der Lieth, C. W. (2007) GlycoMapsDB: A database of the accessible conformational space of glycosidic linkages. *Nucleic Acids Res. 35*, 287−290.
(42) Tvaroška, I., Hricovíni, M., and Petráková, E. (1989) An attempt to derive a new Karplus-type equation of vicinal proton-carbon coupling constants for C-O-C-H segments of bonded atoms. *Carbohydr. Res. 189*, 359−362.
(43) Kitao, A., and Go, N. (1999) Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol. 9*, 164−169.
(44) Gotsev, M. G., and Ivanov, P. M. (2009) Molecular dynamics of large-ring cyclodextrins: Principal component analysis of the conformational interconversions. *J. Phys. Chem. B 113*, 5752−5759.
(45) Poppe, L., Stuike-Prill, R., Meyer, B., and van Halbeek, H. (1992) The solution conformation of sialyl-α (2−6)-lactose studied by modern NMR techniques and Monte Carlo simulations. *J. Biomol. NMR 2*, 109−136.
(46) Liu, J., Stevens, D. J., Haire, L. F., Walker, P. A., Coombs, P. J., Russell, R. J., Gamblin, S. J., and Skehel, J. J. (2009) Structures of receptor complexes formed by hemagglutinins from the Asian Influenza pandemic of 1957. *Proc. Natl. Acad. Sci. U.S.A. 106*, 17175−17180.

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Open Access on 05/30/2015

Insights into the Human Glycan Receptor Conformation of 1918 Pandemic Hemagglutinin−Glycan Complexes Derived from Nuclear Magnetic Resonance and Molecular Dynamics Studies

Stefano Elli,[†] Eleonora Macchi,[†] Timothy R. Rudd,[‡] Rahul Raman,[§] Guillherme Sassaki,[⊥] Karthik Viswanathan,[§] Edwin A. Yates,[∥] Zachary Shriver,[§] Annamaria Naggi,[†] Giangiacomo Torri,[†] Ram Sasisekharan,[*,§] and Marco Guerrini[*,†]

[†]Istituto di Ricerche Chimiche e Biochimiche "G. Ronzoni", Via Giuseppe Colombo 81, 20133 Milano, Italy

[‡]Diamond Light Source Ltd., Diamond House, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, U.K.

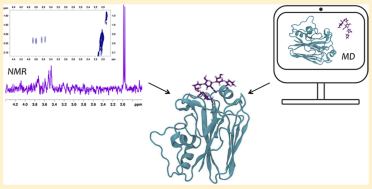
[§]Department of Biological Engineering, Koch Institute of Integrative Cancer Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States

[⊥]Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, CEP 81.531-980, CP 19046, Curitiba, PR, Brazil

[∥]Department of Structural and Chemical Biology, University of Liverpool, Liverpool L69 3BX, U.K.

S *Supporting Information*

ABSTRACT: The glycan receptor binding and specificity of influenza A viral hemagglutinin (HA) are critical for virus infection and transmission in humans. However, ambiguities in the interpretation of the receptor binding specificity of hemagglutinin from human- and avian-adapted viruses have prevented an understanding of its relationship with aerosol transmissibility, an exclusive property of human-adapted viruses. A previous conformational study, which we performed, indicated that human and avian receptors sample distinct conformations in solution. On the basis of detailed nuclear magnetic resonance (NMR) studies provided herein, we offer evidence of the distinct structural constraints imposed by hemagglutinin receptor binding sites on the glycan conformational space upon binding. The hemagglutinin from the SC18 virus, which has efficient aerosol transmissibility in humans (human-adapted), imposed the most stringent constraints on the conformational space of the human glycan receptor (LSTc), compared to single (NY18) or double (AV18) amino acid HA mutants, a property correlating to the ligand−HA binding strength. This relationship was also observed for the avian-adapted HA, where the high affinity binding partner, AV18, imposed the most stringent conformational constraints on the avian receptor, compared to those imposed by NY18. In particular, it is interesting to observe how different HAs when binding to human or avian glycosidic receptors impose significantly different conformational states, in terms of the states sampled by the glycosidic backbone and/or the entire molecule shape (linear or bent), when compared to the corresponding unbound glycans. Significantly, we delineate a "characteristic NMR signature" for the human adapted hemagglutinin (SC18) binding to human glycan receptors. Therefore, the conformational space constraints imposed by the hemagglutinin receptor binding site provide a characteristic signature that could be a useful tool for the surveillance of human adaptation of other (such as H7N9 and H5N1) deadly influenza viruses.



Influenza A viruses are among the most rapidly evolving pathogens with potential for new strains to adapt to human hosts and lead to a pandemic outbreak, with significant economic and public health impact.[1,2] The emergence of novel influenza strains such as 2009 H1N1 and 2010 H3N2 through genetic reassortment[3−5] poses a constant threat in terms of the evolution of various subtypes, including H5N1, H7N2, H7N7, H7N9, and H9N2, to generate a pandemic strain. The H5 and H7 subtypes, among others, are categorized as avian-adapted because they primarily circulate within birds, but through contact with infected animals, they can jump species and infect humans. Some of these avian-adapted viruses, including H5N1 and H7N9, upon infection, can replicate efficiently in various human organs and lead to severe infection and mortality.[6−11] However, these avian-adapted subtypes are not capable of efficient human-to-human aerosol transmission,[12,13] a charac-

Received: March 19, 2014
Revised: May 30, 2014
Published: May 30, 2014

ACS Publications © 2014 American Chemical Society

teristic feature of subtypes such as H1N1, H2N2, and H3N2, which are considered human-adapted.

A key factor governing human adaptation of the influenza A virus is the binding specificity of viral surface glycoprotein hemagglutinin (HA) to sialylated glycan receptors on the host cell surface [glycans terminated by α-D-*N*-acetylneuraminic acid (Neu5Ac)]. A canonical definition of this binding specificity based on the terminal sialic acid linkage has been used in the field in recent decades. HA from avian-adapted subtypes such as H5, H7, and H9 is known to bind specifically to glycans terminated by α(2→3)-linked sialic acid [α(2→3) glycans or avian receptors].[14,15] Meanwhile, HA from human-adapted subtypes such as H1, H2, and H3 is known to bind specifically to glycans terminated by α(2→6)-linked sialic acid.[16,17] This definition based on sialic acid linkage alone, while useful for characterizing many influenza strains, has misclassified some notable strains in terms of their ability to effect efficient human-to-human respiratory droplet transmission,[18−20] in particular H5N1 and H7N9.

Glycan receptors have been defined according to their overall conformation, via a parameter called the θ angle, which defines the form of the nonreducing end of the glycan receptors (Figure 1).[18] Specifically, in the case of influenza, we previously demonstrated that the apical surface of human upper respiratory epithelia, which is a primary target for human-adapted viruses, predominantly expresses glycan receptors with long oligosaccharide branches terminated by α(2→6)-linked sialic acid [α(2→6) glycan or human receptors].[18,21] On the basis of analyses of glycan conformation and topology in HA−glycan X-ray cocrystal structures, we noted that the α(2→6) glycans adopted an "umbrella-like" topology (θ angle of <100°) in the receptor binding site (RBS) of H1 and H2 HAs, while α(2→6) and α(2→3) glycans adopted "cone-like" topologies (θ angle of >100°) in the RBS of H5 HA.[18] This topology-based definition, in addition to the specific sialic acid linkage [i.e., α(2→3) vs α(2→6)], distinguished HA from human-adapted subtypes binding to human receptors from the HA of avian-adapted viruses binding to avian and human receptors.[22]

The relationship between the glycan receptor specificity of HA and the aerosol transmissibility of the virus in ferrets (a well-established animal model for respiratory droplet transmission) was first demonstrated using the prototypic 1918 H1N1 pandemic HA (A/South Carolina/1/18 or SC18).[23] While SC18 showed efficient aerosol transmission in ferrets, a single amino acid mutation in HA (Asp225 → Gly; numbering based on H3 HA) resulted in a virus, NY18, that exhibited inefficient transmission, and a second further mutation, Asp190 → Glu, resulted in a virus, AV18, that could not be transmitted. We previously demonstrated that SC18 HA bound with high specificity and high affinity to human receptors, while AV18 HA bound with high affinity to avian receptors. NY18 bound to both avian and human receptors with a binding affinity substantially lower than that observed for AV18 and SC18. Despite dramatic differences in the aerosol transmissibility of these three viruses and their glycan receptor binding properties, they showed similar infectivity, replication efficiency, and tissue distribution in ferrets directly inoculated with virus. More recently, this approach was extended to the 2009 H1N1 system, revealing that the receptor specificity affects neither replication nor virulence of this pandemic virus in mice or ferrets, again after intranasal inoculation, but did affect animal-to-animal transmission by respiratory droplets.[24] These studies highlight the significance of HA mutations and glycan receptor binding specificity (given that all other genes among the three viruses are intact) in distinguishing aerosol transmissibility from other phenotypic properties of the virus such as infectivity and replication efficiency.

The dramatic changes in relative glycan receptor binding affinities and aerosol transmissibility resulting from single-amino acid changes to SC18 prompted an investigation into the structural nuances governing HA−glycan interactions. While HA−glycan X-ray cocrystal structures revealed differences in overall glycan topology (cone-like vs umbrella-like), when bound to different HAs, these "static" structures did not entirely capture the restrictions imposed on the conformational space of the glycan receptor by the RBS of different HAs, in moving from the free to HA-bound state.

For what is, to the best of our knowledge the first time, we present solution structures of LSTc [human receptor, Neu5Ac-α(2→6)-Gal-β(1→4)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc] and LSTa [avian receptor, Neu5Ac-α(2→3)-Gal-β(1→3)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc] bound to SC18, NY18, and AV18 HAs, based on comprehensive nuclear magnetic resonance (NMR) [saturation transfer difference spectroscopy (STD) and transferred nuclear Overhauser effect spectroscopy (tr-NOESY)] analyses and molecular dynamics (MD) simulations. The combination of both NMR and molecular dynamic simulations is particularly powerful when investigating protein−carbohydrate interactions, specifically when dealing with the highly mobile carbohydrate ligand.[25] We find that

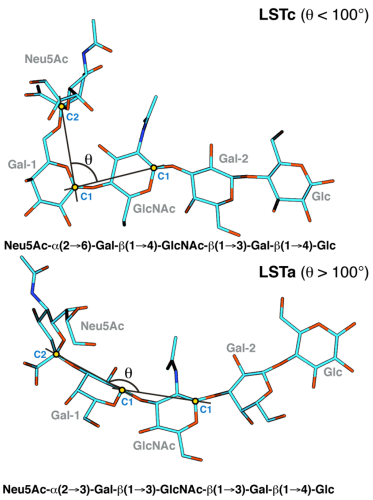


Figure 1. Schematic of LSTc (top) and LSTa (bottom). The θ angle parameter is defined by the angle across anomeric carbons of the three successive residues starting from the nonreducing end (Neu5Ac, Gal-1, and GlcNAc).

these "dynamic" structures are instrumental in delineating the conformational features of glycans when they are bound to HA. Our study captures the active changes in conformation, θ angle, and glycosidic linkage torsional angles of LSTa and LSTc induced upon binding to HA and sheds light on distinct structural constraints imposed by the RBS that differ by one or two amino acids. We report the new finding of defining the restriction on the glycan conformational space and mobility of glycan bound to HA that, importantly, is not captured in the corresponding "snapshot" cocrystal structures. Importantly, these data are consistent with the observed differences in the biochemical binding affinities of these glycans for SC18, NY18, and AV18 HAs. We discuss the implications of these findings for our understanding of binding of HA to human receptors, specifically the significance of glycan conformation, θ angle, and glycosidic torsional angles, and HA RBS interactions in determining the appropriate specificity and affinity. Our findings will greatly aid future studies aimed at delineating appropriate structural constraints on glycan topologies for other influenza subtypes (such as H5N1 and H7N9), including surveillance of adaption to humans by these deadly viruses.

■ MATERIALS AND METHODS

Cloning, Baculovirus Synthesis, Expression, and Purification of HA. The soluble form of HA was expressed using the Baculovirus Expression Vector System (BEVS). SC18 (A/South Carolina/1/1918) baculovirus (generated from the pAcGP67-SC18-HA plasmid[26,27]) was a gift from J. Stevens. pAcGp67-NY18-HA and pAcGp67-AV18-HA plasmids were generated from pAcGP67-SC18-HA by Asp225 → Gly and Asp190 → Glu/Asp225 → Gly mutations, respectively. Mutagenesis was conducted using the QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene). The primers used for mutagenesis were designed using the web-based program PrimerX (http://bioinformatics.org/primerx/) and synthesized by IDT DNA technologies (Coralville, IA). NY18 and AV18 baculoviruses were created from pAcGP67-NY18-HA and pAcGP67-AV18-HA constructs using a Baculogold system (BD Biosciences, San Jose, CA) according to the manufacturer's instructions. The baculoviruses were used to infect 300 mL suspension cultures of Sf9 cells (BD Biosciences) cultured in BD BaculoGold Max-XP Insect Cell medium (BD Biosciences). These cultures were monitored for signs of infection and harvested 4−5 days postinfection. BEVS produces trimeric HA that provides multivalent binding to glycans. The soluble form of HA was purified from the supernatant of infected cells using the protocol described previously.[27] Briefly, the supernatant was concentrated using Centricon Plus-70 centrifugal filters (Millipore, Billerica, MA), and the trimeric HA was recovered from the concentrated cell supernatant using affinity chromatography with columns packed with Ni-NTA beads (Qiagen, Valencia, CA). Eluting fractions that contained HA were pooled and dialyzed overnight with 10 mM Tris-HCl, 50 mM NaCl buffer (pH 8.0). Ion exchange chromatography was then performed on the dialyzed samples using a Mono-Q HR10/10 column (GE healthcare, Piscataway, NJ). The fractions containing HA were pooled together and subjected to ultrafiltration using Amicon Ultra 100 K NMWL membrane filters (Millipore). The protein was then concentrated and reconstituted in PBS. The purified protein was quantified using Bio-Rad's protein assay (Bio-Rad, Hercules, CA).

NMR Analysis of SC18, NY18, and AV18 with LSTc and LSTa. STD and tr-NOESY samples were prepared by washing the proteins SC18, NY18, and AV18 (1 mg/mL) with a buffered solution [150 mM sodium chloride, 100 mM sodium phosphate, 0.3 mM *d*-EDTA, and D_2O (pH 7.2)] using Amicon Ultra centrifugal filters, and a 10 kDa membrane (Millipore). Each ligand (LSTc or LSTa) was added to the corresponding protein sample, yielding final molar ratios of 100:1 (glycan receptor:HA) for the STD measurements and 25:1 (glycan receptor:HA) for the tr-NOESY measurements, and the protein concentration for the STD measurements was 0.01 mM and for the tr-NOESY experiments was 0.04 mM. NMR spectra were acquired using a Bruker 600 and 900 MHz AVANCE series NMR spectrometer, both equipped with a high-sensitivity 5 mm TCI cryoprobe. LSTc and LSTa resonances were previously assigned.[28] For the STD experiments, the on-resonance frequency was set at 7.3 ppm (6600 Hz) and the off-resonance frequency at 20.0 ppm (18000 Hz), a train of 40 Gaussian-shaped pulses of 50 ms each were applied to produce a selective saturation of 2 s, and D1 was 6 s. The number of scans was 1K, and the spectral width was 12626 Hz. The two-dimensional NOESY experiments were conducted using a mixing time of 300 ms; the data were recorded for quadrature detection in the indirect dimension and acquired using 16 scans per series of 2048 × 416 data points. The spectra were recorded at 295 K.

Dose-Dependent Direct Binding of SC18, NY18, and AV18 to Glycan Receptors LSTa and LSTc. LS-tetrasaccharide c [LSTc, Neu5Ac-α(2→6)-Gal-β(1→4)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc)] and LS-tetrasaccharide a [LSTa, Neu5Ac-α(2→3)-Gal-β(1→3)-GlcNAc-β(1→3)-Gal-β(1→4)-Glc] (Accurate Chemicals) were biotinylated with EZ-Link Biotin-LC-Hydrazide (Thermo) according to the manufacturer's instructions. Streptavidin-coated High Binding Capacity 384-well plates (Pierce) were loaded to the full capacity of each well by incubating the well with 50 μL of 2.4 μM biotinylated LSTa or LSTc overnight at 4 °C. Excess glycans were removed through extensive washing with PBS. The trimeric HA unit comprises three HA monomers, and the spatial arrangement of the biotinylated glycans in the wells of the streptavidin plate array favors binding to only one of the three HA monomers in the trimeric HA unit. To specifically enhance the correct multivalency in the HA−glycan interactions, the recombinant HA proteins were precomplexed with the primary and secondary antibodies in a molar ratio of 4:2:1 (HA:primary:secondary). The identical arrangement of four trimeric HA units in the precomplex for all the HAs permitted comparison between their glycan binding affinities. A stock solution containing appropriate amounts of histidine-tagged HA protein, primary antibody (mouse anti-six-His tag IgG), and secondary antibody (HRP-conjugated goat anti-mouse IgG) (Santa Cruz Biotechnology, Santa Cruz, CA) was combined in a ratio 4:2:1 and incubated on ice for 20 min. Appropriate amounts of precomplexed stock HA were diluted to 250 μL with 1% BSA in PBS; 50 μL of this precomplexed HA was added to each of the glycan-coated wells and incubated at room temperature for 2 h followed by the wash steps described above. The binding signal was determined on the basis of HRP activity using the Amplex Red Peroxidase Assay (Invitrogen, CA) according to the manufacturer's instructions. The experiments were conducted in triplicate. Minimal binding signals were observed in the negative controls, including binding of the precomplexed unit to wells without glycans and binding of the antibodies alone to the wells with glycans. The

data obtained from this analysis can be found in Figure 1 of the Supporting Information.

Molecular Dynamics Simulations. The dynamic and conformational properties of the interaction between LSTc and HA were studied by comparing the MD simulation trajectories of three complexes: LSTc−SC18, LSTc−NY18, and LSTc−AV18. The X-ray cocrystal structures of SC18 and NY18 were recently determined with LSTa and LSTc.[29] However, during the preparation of this paper, the available cocrystal structures were those of SC18 with LSTc [Protein Data Bank (PDB) entry 2WRG], a swine H1N1 HA (A/swine/Iowa/30) with LSTa and LSTc, and another human H1N1 (A/Puerto Rico/8/34, PDB entry 1RVX[30]) with LSTa. These X-ray cocrystal structures constituted reasonable starting models of HA−glycan complexes for MD simulations. The LSTc−SC18 complex was built starting with LSTc cocrystallized with SC18 HA [PDB entry 2WRG; coordinates were available for tetrasaccharide Neu5Ac-α(2→6)-Gal-β(1→4)-GlcNAc-β(1→3)-Gal-].[31] The protein, HA portion, of the complex was taken directly from the PDB coordinates, selecting a sequence of 60−260 amino acids (2WRG numbering) that includes the HA RBS. The solution conformation of LSTc, determined previously,[28] was superposed onto the cocrystallized glycan structure, with the nonreducing residues Neu5Ac and Gal-1 giving a root-mean-square deviation (rmsd) of 6.5 Å. The remaining complexes were obtained from the previously built LSTc−SC18 complex by applying the single mutation Asp225 → Gly *in silico* to generate the LSTc−NY18 complex and the double mutation Asp225 → Gly/Asp190 → Glu to generate the LSTc−AV18 complex. It is important to note that at the start of the MD simulation, the three complexes are characterized by exactly the same conformation of LSTc and HA with the exception of mutated residues.

The LSTa−AV18 complex was built by superimposing a previously selected conformation of LSTa[28] on the LSTa-like trisaccharides cocrystallized with H1 (A/Puerto Rico/8/34 H1N1) with amino acids within the binding site typical of an avian HA (PDB entry 1RVX). The rmsd between the superposed glycans is 6.2 Å, calculated on residues Neu5Ac and Gal-1. The previously built AV18 protein was superposed on the HA in PDB entry 1RVX, matching the protein Cα backbone (rmsd = 0.38 Å); then the complex was built by taking LSTa and AV18. The LSTa−NY18 complex was built from the latter complex (LSTa−AV18) by substituting Glu190 with Asp. Even in that case, the LSTa−AV18 and LSTa−NY18 model complexes have the same geometry at the beginning of the MD simulations.

Ambertools 1.4 was used to build the force field. GLYCAM06/Amber was used to describe the glycan and protein part of the complexes. The simulation cell was built by enveloping each macromolecule by a water layer (TIP3P water model) 15 Å wide in three directions, resulting in an orthogonal cell with an edge of approximately 100 Å. The nonbonded potential energy was described using the standard cutoff (12 Å) technique for both electrostatic and dispersive interactions. Each cell was minimized using 100 K steps of the default minimization algorithm included in NAMD. Then 1 ns of MD simulation sampling the *NPT* ensemble was used to equilibrate the cell density. The simulation temperature was set at 295 K and maintained by a Langevin thermostat as implemented in NAMD, while the Nosé-Hoover Langevin piston algorithm controlled the pressure (1.01325 bar) applied to the cell walls. During the minimization and cell density equilibration steps, a harmonic potential energy restraint (harmonic constant of 50 kcal mol^{-1}) on all the atoms of the complex was applied, while the water molecules were allowed to move freely. The MD simulation for all the modeled complexes was ∼120 ns and was completed by applying a soft harmonic restraint on the HA backbone atoms (Cα, N, and carbonyl carbon) with a harmonic constant of 2.0 kcal mol^{-1}. This allows the ligand and the side chain residues to be adjusted, while the secondary structure elements are maintained. The MD simulation trajectory was sampled every 10 ps, and the comparisons between the different complexes were conducted by monitoring selected distances between the ligand and the HA active site residues (Figure 4).

MD Simulation Trajectory Principal Component Analysis (PCA). PCA has been a powerful tool for the analysis of protein[32,33] and glycan[28,34] molecular dynamics. Although in these examples the PCA was used with the aim of extracting distinct and independent motional modes, it has been used in our study to identify the final bound states in the LSTx−HA complexes. Each frame of the glycan−HA MD trajectories was converted from Cartesian coordinates to a distance matrix, measured between the glycan and protein. The distances were determined between the non-carbon and hydrogen atoms of the glycan (excluding the glycosidic linkage oxygens and including the carbon of the *N*-acetyl groups) and the amino acid side chains of HA, the last carbon in the amino acid side chain. A 6 Å cutoff was applied to the distance matrices; this means that only glycan−HA interactions were observed and not the glycan rearranging in solution, away from the RBS. These matrices were then converted into a vector, and all of them, for a single MD trajectory, were placed into a matrix; this data set was then mean-centered before principal component analysis could be performed. Density-based cluster analysis was performed on the first two component loadings (the most significant); the time each cluster appeared in the MD trajectory could then be compared. The first 20 ns of each trajectory was discarded; this left the time interval from 20 to 120 ns to be investigated, which was decomposed into 10000 distance matrices that were examined by PCA. This approach allows the evolution of the glycan−HA complex to be observed, from the initial state to the final state. For the final bound states, the ϕ, ψ, and θ angles were determined as well as the average glycan−protein contacts for that subset of conformers. The glycan−protein contacts are represented as networks, with the edge thickness being inversely proportional to distance (the thicker the edge, the closer the vertices are); these can be found in Figure 5 and Figures 8−11 of the Supporting Information.

Parameters That Define the Glycan Conformation and Topology. The torsional angles (ϕ and ψ) are defined as the following pairs of dihedral angles: ϕ_1 and ψ_1, ϕ_2 and ψ_2, ϕ_3 and ψ_3, and ϕ_4 and ψ_4 (starting from the nonreducing termini). For LSTa, the first pair is defined as C1−C2−O3−C3 (ϕ_1) and C2−O3−C3−H3 (ψ_1), while for LSTc, ϕ_1 is the C1−C2−O6−C6 angle and ψ_1 the C2−O6−C6−C5 angle, as previously defined by Xu et al.[35] Thereafter, successive pairs are defined as H1−C1−O4′−C4′ (ϕ_i) and C1−O4′-C4′−H4′ (ψ_i) for the 1→4 linkage or H1−C1−O3′−C3′ (ϕ_i) and C1−O3′−C3′−H3′ (ψ_i) for the 1→3 linkage. Atoms labeled with a prime belong to the monosaccharide on the reducing side of the glycosidic linkage, while atoms without a prime are on the nonreducing side of the glycosidic linkage. To be consistent with Chandrasekaran et al.,[18] Xu et al.,[35] and Sassaki et al.,[28] torsional angles ϕ and ψ were illustrated in the range of −120°

Table 1. Glycosidic Torsional Angles for LSTc, Free and Bound to SC18, NY18, and AV18[a]

	linkage	cluster	size	percentage	ϕ (deg)	standard deviation (SD) of ϕ (deg)	ψ (deg)	SD of ψ (deg)
LSTc	1	1	9858	98.6	−57	11	190	17
		2	48	0.5	−71	12	−114	5
		3	73	0.7	−60	9	111	13
	2	1	5430	54.3	42	11	−3	11
		2	2419	24.2	−38	13	−29	11
		3	2151	21.5	24	13	−32	10
	3	1	4691	46.9	20	14	−47	13
		2	3981	39.8	39	15	34	18
		3	1328	13.3	−25	11	−43	11
	4	1	5934	59.3	39	13	−8	18
		2	4066	40.7	−37	14	−30	13
LSTc–SC18	1	1	2763	100.0	−58	7	189	7
	2	1	2763	100.0	50	9	−5	9
	3	1	2724	98.6	−2	21	−42	11
		2	39	1.4	40	7	25	8
	4	1	1192	43.1	45	10	0	11
		2	1069	38.7	19	13	−30	11
		3	502	18.2	−33	18	−28	11
LSTc–NY18	1	1	1826	100.0	−46	10	194	7
	2	1	1826	100.0	41	9	−7	11
	3	1	1826	100.0	17	14	−43	12
	4	1	1826	100.0	−31	13	−27	9
LSTc–AV18	1	1	2698	100.0	−51	8	192	7
	2	1	2698	100.0	−24	14	−32	11
	3	1	2698	100.0	24	13	34	11
	4	1	2698	100.0	−28	12	−24	9

[a]These angles for the ligand–receptor complexes were determined for those conformers that were extracted by PCA: g3, LSTc–SC18; g2, LSTc–NY18; and g2, LSTc–AV18 (Figure 6 of the Supporting Information). The angles were determined by cluster analysis of the data illustrated in Figure 13 of the Supporting Information, and the analysis used nonparametric density estimation[36] to determine the members of each cluster. These angles are average values for each ϕ and ψ cluster.

to 240°. The torsional angles for the LSTc–HA and LSTa–HA complexes are listed in Tables 1 and 2, respectively. The angles were determined by cluster analysis of the data illustrated in Figures 13 and 14 of the Supporting Information, and the analysis used nonparametric density estimation[36] to determine the members of each cluster. The angles represented in the tables are the average values for each ϕ and ψ cluster.

The topology of LSTa and LSTc is defined using a θ angle parameter. The θ angle is defined by the C2, C1, and C1 atoms of residues Neu5Ac, Gal, and GlcNAc (*N*-acetylglucosamine), going from the nonreducing end to the reducing end (Figure 1).

RESULTS

NMR Structural Analyses of HA–Glycan Interactions. NMR analyses were performed using a 900 MHz spectrometer, which permitted unprecedented resolution for observation of HA–glycan interactions. NMR STD experiments were employed to obtain a qualitative description of the glycan residues that interact with HA. NMR analysis indicated that for both LSTa and LSTc (Figure 1) in complex with the different HAs, the main sugar residue involved in the contact is the terminal sialic acid (Neu5Ac). However, there are differences in the mode of binding for these glycans to the various HAs.

To aid the assignment of the one-dimensional STD signals, spectra of LSTc bound to SC18 and NY18 were superimposed on the HSQC spectrum of LSTc (Figures 2A and 3A). While the interaction between LSTc and SC18 occurs primarily through the nonreducing end Neu5Ac moiety, involving protons H3ax, H3eq, H4, H5, H7, and H9 (Figure 2A), signals belonging to Gal-1 (H6), GlcNAc (methyl group), Gal-2 (H4 and H1), and Glc (H5 and H6) are also present in the STD spectrum (Figure 2A). Similarly, LSTc interacts with NY18 principally through Neu5Ac, (H4, H5, H7, and H9), whereas signals belonging to Gal-2 and Glc appear to be absent or, at best, weaker than those found in the STD spectrum of the LSTc–SC18 complex. Notably, both methyl groups of GlcNAc and Neu5Ac still interact with NY18, but these signals are weaker than those observed in the LSTc–SC18 complex (Figure 3A).

The greater number of proton signals corresponding to the monosaccharides of LSTc that interact with SC18 compared to NY18 is consistent with the dose-dependent direct biochemical binding of these HAs to this glycan in an array platform (Figure 1 of the Supporting Information) and also with previously reported human receptor affinity of SC18 HA that is higher than that of NY18 HA.[20] The absence of STD signals observed for the LSTc–AV18 complex (Figure 3 of the Supporting Information) indicates that LSTc does not interact with AV18, consistent with minimal to no binding of AV18 to LSTc that is observed in the dose-dependent binding assay (Figure 1 of the Supporting Information).

The avian receptor, LSTa, on the other hand, interacts with AV18 and NY18 almost exclusively through Neu5Ac (Figure 2B). Because of the overlapping signals of the methyl groups belonging to Neu5Ac and GlcNAc, it was not possible to establish definitively which group interacts with the corresponding HA (Figure 3B). Among the few STD signals that

Table 2. Glycosidic Torsional Angles for LSTa, Free and Bound to NY18 and AV18[a]

	linkage	cluster	size	percentage	ϕ (deg)	SD of ϕ (deg)	ψ (deg)	SD of ψ (deg)
LSTa	1	1	8771	87.7	−62	10	−7	13
		2	1191	11.9	−81	9	−55	9
		3	21	0.2	230	6	47	8
		4	17	0.2	209	12	−26	8
	2	1	7317	73.2	41	13	−1	17
		2	1917	19.2	−37	13	−20	13
		3	766	7.7	61	12	57	10
	3	1	5939	59.4	39	16	32	19
		2	3290	32.9	20	13	−48	14
		3	771	7.7	−28	11	−44	13
	4	1	7861	78.6	40	13	−6	17
		2	2139	21.4	−34	14	−30	12
LSTa–AV18	1	1	2426	100	206	15	−20	8
	2	1	2426	100	51	9	19	15
	3	1	1118	46.1	49	9	53	13
		2	1308	53.9	25	12	42	12
	4	1	2426	100	54	22	5	21
LSTa–NY18	1	1	3421	97.1	−64	11	−6	14
		2	101	2.9	−73	10	−50	8
	2	1	2674	75.9	−40	11	−16	14
		2	484	13.7	45	11	1	12
		3	364	10.3	23	13	−26	10
	3	1	1939	55.1	22	15	−44	14
		2	1234	35	−24	17	−42	14
		3	349	9.9	47	14	14	19
	4	1	1760	50	−36	14	−30	12
		2	1762	50	35	15	−11	19

[a]These angles for the ligand–receptor complexes were determined for those conformers that were extracted by PCA: g2–g5, LSTa–NY18; g1 and g2, LSTa–AV18 (Figure 7 of the Supporting Information). Unlike when LSTc is bound to HA, multiple bound states were evident when LSTa was bound to HA. These angles were determined by cluster analysis of the data illustrated in Figure 14 of the Supporting Information, and the analysis used nonparametric density estimation[36] to determine the members of each cluster. These angles are average values for each ϕ and ψ cluster.

were observed for the interaction between LSTa and AV18 or NY18 that did not belong to Neu5Ac, weak signals were observed for H3 and H4 of Gal-1, consistent with the partial involvement of the Gal-1 residue in binding.

Given the conformational flexibility of glycans, arising from the numerous glycosidic torsion angles, the parameter θ has been defined as quantifying the form of the nonreducing end of the receptors (Figure 1 and Materials and Methods). To compare the conformational space sampled by the free and bound ligand, tr-NOESY experiments were performed for LSTc interacting with SC18 and NY18 (Figure 5 of the Supporting Information). Notably, via comparison of the results from bound and unbound LSTc, the NOE signals of the GlcNAc methyl group are substantially different. Whereas only the NOE between the methyl group of GlcNAc and H5 of Neu5Ac was observed for unbound LSTc, additional NOEs are observed in the bound state with SC18, including signals associated with protons H8 and H9. These findings agree with a decrease in the distance between the GlcNAc (CH_3 protons) and Neu5Ac (H5 and H8/H9 protons), which is probably correlated with a reduction in the θ angle between the free and bound state. This analysis indicates that, upon binding to SC18 and to a lesser extent NY18, LSTc undergoes a conformational change, reducing the θ angle and subsequently the level of conformational freedom of the Neu5Ac residue, particularly in the region between atoms C6 and C9.

Molecular Dynamics Simulation of HA–Glycan Interactions. The impact of glycan conformational differences and points of glycan–HA interactions observed in the NMR experiments was further investigated using MD simulations, allowing a structural and dynamic comparison between the various complexes (LSTc–SC18, –NY18, and –AV18 and LSTa–AV18 and –NY18) to be made.

During the MD simulations, the conformational and dynamic properties of the HA–glycan complexes progressively change, especially those of the glycan conformation, with differences arising in the complexes due to the amino acid mutations within the RBS. It should be noted that the glycan starting geometries are the same in the two sets of MD simulations, with the LSTa and LSTc starting geometries taken from the solution forms identified by Sassaki et al.[28] The mobility observed for the glycan in the HA RBS was monitored using PCA, as described in Materials and Methods. Briefly, the analysis was performed on the distance matrix between the HA and receptor; unlike the conventional PCA of protein MD simulations, which uses the position of the protein backbone or the dihedral angles of the protein, this focuses the analysis on the interaction between the receptor and ligand. The temporal changes in the conformational state of the LSTc–SC18 complex are reported in Figure 6C of the Supporting Information; the conformer subset identified by cluster g3 represents the final bound state of the complex, and a similar representation for the LSTc–NY18 and LSTc–AV18 complexes is shown in Figure 6 of the Supporting Information. When LSTc is bound to any of the HAs, it appears to find a final singular state; this is not the case for the LSTa–AV18 or

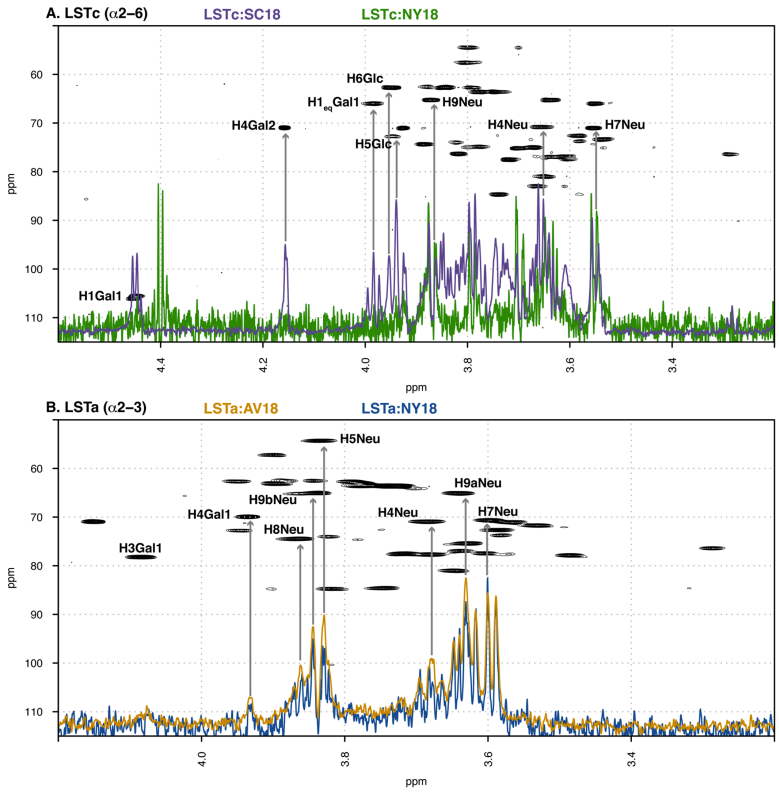


Figure 2. Main regions of the overlaid STD HSQC spectra of LSTc– and LSTa–receptor complexes. (A) STD spectra of LSTc–SC18 (purple) and LSTc–NY18 (green) complexes overlaid upon the HSQC spectrum of LSTc. (B) STD spectra of LSTa–AV18 (orange) and LSTa–NY18 (blue) complexes overlaid upon the HSQC spectrum of LSTa. The HSQC spectra and chemical shift assignments of LSTc and LSTa can be found in ref 28. Individual figures, with each STD spectrum plotted over the relevant glycan HSQC spectrum, can be found in Figures 2 and 4 of the Supporting Information.

LSTa–NY18 complex, where the final state of the latter MD simulation is represented by at least four conformational subsets (Figure 7C of the Supporting Information, clusters g2–g5) and the former has two final conformational subsets (Figure 7F of the Supporting Information, clusters g1 and g2), indicating a greater level of conformational freedom compared to that for the case in which LSTc interacts with HA.

Inspection of the MD trajectory indicates that, in the case of the LSTc–SC18 complex, all the monosaccharides of LSTc are positioned to interact with the RBS (Figure 5A and Table 1 of the Supporting Information). It is also apparent that both hydrogen bonds and dispersive forces are important components in the interaction of LSTc with SC18; specifically, the methyl groups of GlcNAc show persistent contacts with Leu194 and Asp190, while the methyl group of Neu5Ac interacts with Gly134 and Trp153 (Figure 5A). Other noteworthy interactions within the LSTc–SC18 complex are between Gal-1 and Gln226, Lys222, and Asp225.

The presence of Gly instead of Asp at position 225, in going from wild-type SC18 to NY18, removes the hydrogen bond

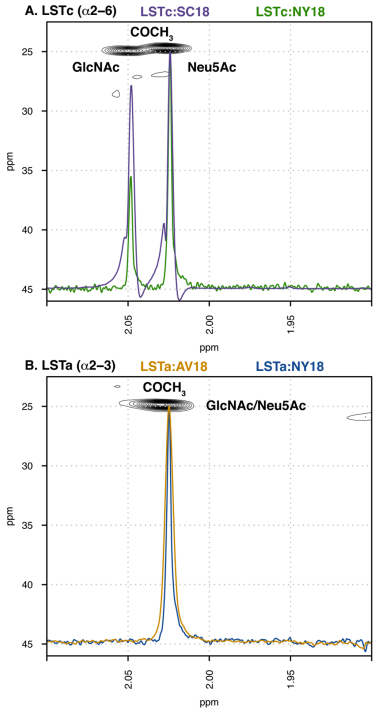


Figure 3. *N*-Acetyl regions of the overlaid STD HSQC spectra of LSTc– and LSTa–receptor complexes. (A) STD spectra of LSTc–SC18 (purple) and LSTc–NY18 (green) complexes overlaid upon the HSQC spectrum of LSTc. (B) STD spectra of LSTa–AV18 (orange) and LSTa–NY18 (blue) complexes overlaid upon the HSQC spectrum of LSTa. The HSQC spectra and chemical shift assignments of LSTc and LSTa can be found in ref 28. Individual figures, with each STD spectrum plotted over the relevant glycan HSQC spectrum, can be found in Figures 2 and 4 of the Supporting Information.

interaction between Gal-1 (OH3 and OH4) and the RBS (specifically involving residues Gln226, Lys222, and Asp225), thereby promoting a greater distance between Gal-1 and HA (loop220) compared to that in SC18 (Figures 3B and 5B and Table 1 of the Supporting Information). This result is clearly supported by the NMR STD data (Figures 2 and 3), where Gal-1, GlcNAc, Gal-2, and Glc resonances are absent from the spectrum of the LSTc–NY18 complex or weaker than those in the spectrum of the wild-type complex, corresponding to a weaker interaction.

In the LSTc–AV18 complex, the further mutation of Asp190 to Glu introduces greater steric hindrance by preventing the optimal interaction between the reducing end of LSTc and helix190 of AV18, as can be seen by comparing the corresponding distances in the MD simulation trajectories of the LSTc–AV18 complex with those of the LSTc–SC18 and LSTc–NY18 reference complexes (Figure 5 and Table 1 of the Supporting Information). The consequence of this is that the interaction between Gal-1 and the HA RBS is re-established and Neu5Ac is drawn closer to the RBS, while the interaction of GlcNAc is weaker than that in the LSTc–SC18 and LSTc–NY18 complexes (specifically, Asp190 and Leu194 with GlcNAc-CMe), with the whole residue moving away from the RBS.

The analyses here reaffirm that it is interaction of Neu5Ac, Gal-1, and GlcNAc with HA that is important for human adaptation of influenza. These results provide a structural description of the effect caused by the single- and double-amino acid mutations in the RBS of HA, which correlates with a progressive weakening of the interaction between LSTc and HA (SC18 > NY18 > AV18), a result also supported by NMR STD experiments and the measured binding affinities.

In the case of the LSTa–AV18 complex, the primary contacts involve Neu5Ac of LSTa, with the carboxyl, acetyl, and sialyl groups of Neu5Ac interacting with the HA RBS (Figure 4C and Figure 8 and Table 2 of the Supporting Information). Importantly, contacts are also observed between AV18 and Gal-1 in LSTa. This involves the interaction of O6-Gal-2 with Glu190 and Pro186, which is afforded by the presence of the $\alpha(2\rightarrow3)$ linkage in LSTa. PCA of the LSTa–AV18 MD simulation trajectory also indicates the mobility of the reducing end region of LSTa is greater than that of its nonreducing terminus (Neu5Ac), which remains strongly attached to the RBS on the MD simulation time scale (Figures 7 and Figure 8 and Table 2 of the Supporting Information). In the case of the LSTa–NY18 complex, the mutation of Glu190 to Asp reduces the extent of interaction between Gal-1 and the RBS, with only the nonreducing end Neu5Ac residue interacting with the RBS. A consequence of this is that the number of interactions between the sialyl group of Neu5Ac and the RBS is decreased and the interaction between Gln226 and the carboxyl group of Neu5Ac is no longer observed (Figure 4D and Figure 9 and Table 3 of the Supporting Information). This corresponds to a weaker interaction between LSTa and NY18 than between LSTa and AV18, in agreement with the NMR STD experiments and biochemical assay. The LSTa–NY18 complex samples two conformational states during the MD simulation, which differ on the basis of contacts between Neu5Ac and the HA RBS (Figure 9 of the Supporting Information). The features of binding of LSTa to AV18 and NY18 obtained from the MD simulations are consistent with the corresponding NMR STD signals, where data from both complexes indicate that Neu5Ac is the main interacting residue with HA, while the methyl group of GlcNAc is no longer a key point of interaction with either NY18 or AV18.

We have compared the available X-ray crystallographic structures with our MD and NMR structures (Figures 10 and 11 of the Supporting Information). The comparisons have been limited to glycan–protein contact networks as the forms of the glycans within the crystal structures are distorted, precluding the determination of conformational angles.

Dynamics of LSTa and LSTc Conformations upon Binding to HA. The θ angle parameter is a key conformational

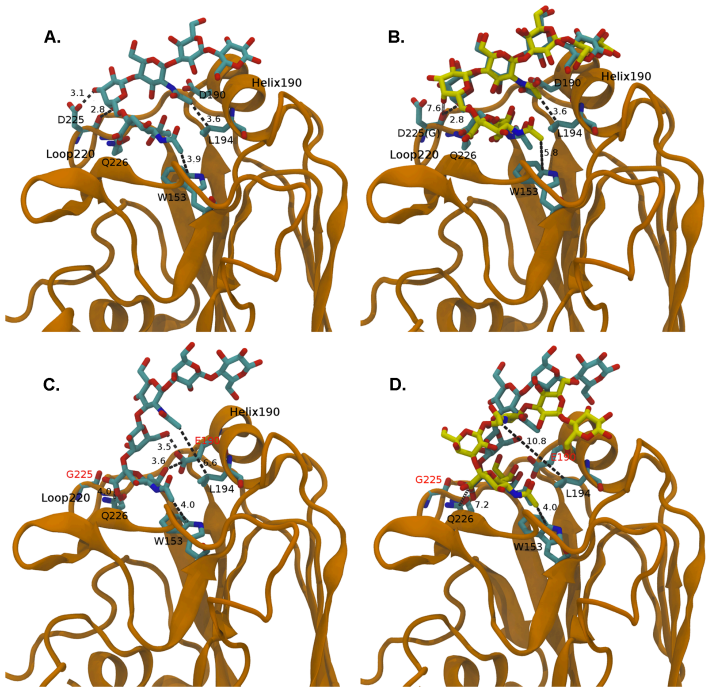


Figure 4. (A) Structure of the LSTc−SC18 complex in the g3 cluster conformation subset (Figure 6C of the Supporting Information). (B) Superposition of the RBSs of the LSTc−SC18 complex with that of the LSTc−NY18 complex (rmsd = 0.78 Å). The reported complex structures belong to g3 and g2 conformations, respectively (Figure 6C,F of the Supporting Information). The carbon skeleton of LSTc bound to SC18 is colored cyan, while LSTc interacting with NY18 is colored yellow. (C) Structure of the LSTa−AV18 complex corresponding to conformation subset g2 of the PCA conformational characterization (Figure 7C of the Supporting Information). (D) Superimposed structures of the LSTa−AV18 and LSTa−NY18 complexes (rmsd = 0.74 Å). The reported structure of the LSTa−NY18 complex corresponds to g2 and g1 conformation subsets as obtained from PCA (Figure 7F of the Supporting Information). LSTa linked to AV18 is colored cyan, while LSTa interacting with NY18 is colored yellow. The relevant amino acid residues of the shown HA active sites are underlined by a tube representation, with the name and numbering relative to PDB entry 2WRG. The reported distances are in angstroms.

descriptor of the nonreducing end of the glycan receptor, indicating the different forms of the glycans. As shown previously, the θ angle of unbound LSTc had a predominant distribution at 86° and a smaller population located at 119° (Figure 6A).[28] Binding to SC18 substantially restricts the conformational population of LSTc, as reflected by the narrow distribution of the θ parameter, with a single distribution centered at 82°. Binding to NY18 also restricts the conformational space sampled by LSTc, with a θ angle distribution being located at approximately 90°. Interestingly, the small population of θ angle values around 119° in the unbound LSTc is absent in the bound state, consistent with an earlier study in which it was postulated that long $\alpha(2\rightarrow6)$ glycans would predominantly adopt an umbrella-like topology (characterized by $\theta < 100°$) when bound to the RBS of human-adapted HAs.[18]

This restriction of the θ angle when LSTc binds to SC18 and NY18 is supported experimentally by the new NOE signal appearing between GlcNAc and Neu5Ac H8/H9 in the tr-NOESY spectra of the LSTc−SC18 complex and by the stronger NOE signal between GlcNAc and Neu5Ac H5 in the LSTc−NY18 complex (Figure 5 of the Supporting Information). In contrast, low-affinity interaction of LSTc with AV18 predominantly samples the cone-like topology as indicated by

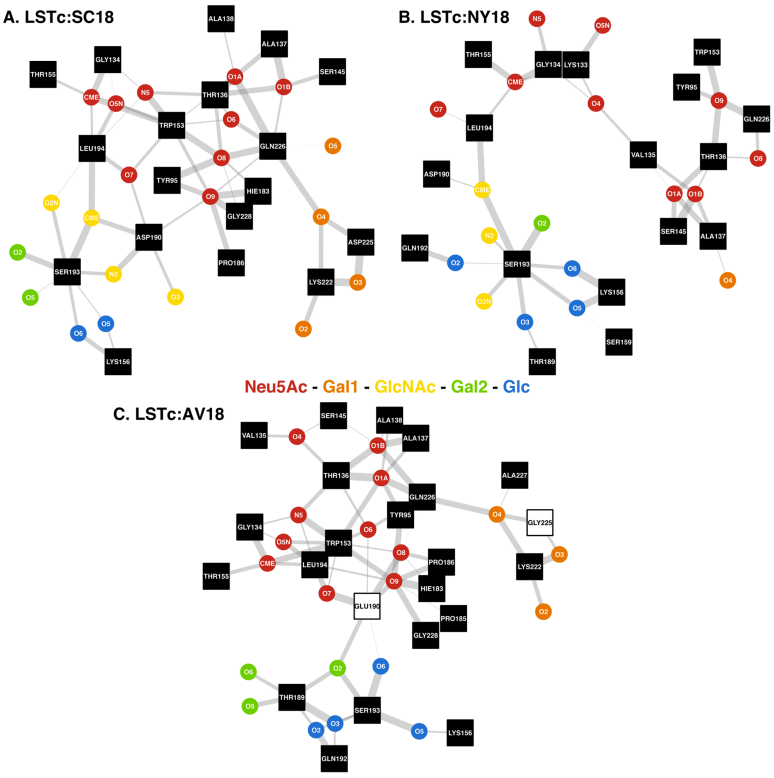


Figure 5. Contact network between LSTc and HA. If two vertices are linked, they are <6 Å apart; the thicknesses of the graph edges are inversely proportional to the distance between the glycan atom (circular vertex) and the protein amino acid (square vertex) (the thicker the edge, the closer the two are together). The square vertices that are white with a black boundary indicate that this amino acid has been mutated, for example, Asp190 → Glu. The networks represent the average distances found in the final, "bound", conformer subsets identified by PCA (Figure 6 of the Supporting Information): cluster g3, LSTc−SC18; cluster g2, LSTc−AV18; and cluster g2, LSTc−NY18. The distances illustrated here are listed in Table 1 of the Supporting Information. In this figure, the distinct modes of the interaction between LSTc and HA (SC18, NY18, and AV18) can be observed. When binding to its natural ligand, SC18, LSTc interacts with HA along its entire length. The single-point mutation forming NY18, Asp225 → Gly, drastically alters the interaction between the nonreducing end of the receptor and HA, and the major interactions with Gal-1 are abolished (panel B compared to panel A). The additional modification forming AV18, Asp190 → Glu, allows Gal-1 to re-engage with HA, which leads to the interaction between GlcNAc and amino acids 190 and Ser193 being lost (compare panel C to panels B and A).

the θ distribution around 113°, which is consistent with the constraints imposed by avian-adapted HAs[18] and is due to the lack on interaction between GlcNAc and the RBS. These results re-emphasize the ability of human receptors to sample a distinct set of topologies (both umbrella-like and cone-like) in the RBS of SC18, NY18, and AV18 HA.

In the unbound state, the θ angle of LSTa samples two distinct populations centered around 118° and 154° (Figure 6B), both of which correspond to cone-like topologies (θ > 100°). The binding of LSTa to NY18 does not significantly alter the position of the θ angle distribution compared to that of free glycan, distributions located at 117° and 158° in the

bound state, but does affect strongly their relative populations, with the population located at the smaller θ angle being the largest. The binding of LSTa to AV18 HA imposes restrictions on the θ angle, removing the population distributed around 117° (Figure 6B). Consistent with our previous studies, the avian receptor LSTa adopts exclusively a cone-like topology regardless of whether it is present in the unbound form or is bound to NY18 or AV18 HA. For the sake of completeness, the θ angle density plots for the entire MD trajectory (20−100 ns) can be found in Figure 12 of the Supporting Information.

To complete our conformational studies, we also considered the glycosidic torsion angles of the free and bound glycan receptor sampled during the MD simulation; in the case of the bound receptors, the conformer subset extracted by PCA is considered (Tables 1 and 2).

First, consistent with the θ angle distributions described above, comparison of the glycosidic torsional angles in the unbound and various HA-bound glycans indicates there are distinct structural constraints imposed by the RBS of SC18, NY18, and AV18 HA on LSTa and LSTc. When LSTc binds to SC18 and NY18, the distributions of states seen for ϕ_1 and ψ_1, ϕ_2 and ψ_2, and ϕ_3 and ψ_3 are very similar; only by using density cluster analysis is it possible to observe the different central locations of the states (Figure 13 and Table 1 of the Supporting Information). Indicating the modification of amino acid 190 (Asp → Gly) and the subsequent change in the interaction between LSTc and HA, the loss of the interaction of Gal-1 with Gln226, Asp225, and Lys222 does not have a strong effect on the glycosidic torsional angles, whereas binding of LSTc to AV18 modifies amino acids 190 and 225, which eradicates the interaction between GlcNAc and Leu194, Asp190, and Ser193, affecting the positions of ϕ_2 and ψ_2 and those of ϕ_3 and ψ_3, which are distinct from those observed in the LSTc−SC18 and LSTc−NY18 complexes (Figure 13 and Table 1 of the Supporting Information).

When LSTa binds to HA, the interactions with AV18 and NY18 are very different. The interaction between LSTa and NY18 is solely through the nonreducing end Neu5Ac (Figure 9 of the Supporting Information), and this is evident in the diversity of states observed for ϕ_2 and ψ_2, ϕ_3 and ψ_3, and ϕ_4 and ψ_4 (Figure 14 and Table 2 of the Supporting Information). The opposite is seen for the LSTa−AV18 complex; even though the glycan is dynamic with four conformational states being extracted by PCA of the MD trajectory of the complexes (Figure 7 of the Supporting Information), the glycosidic torsional angle states observed in the glycan are restricted for all four linkages.

■ DISCUSSION AND CONCLUSION

The NMR and MD simulation analyses described in this study offer new insights into the interaction between hemagglutinin and its glycan receptors, providing a detailed description of the contacts observed in the interactions between LSTc and LSTa with human- and avian-adapted HA and the consequent change in glycan conformation. The principal consequence of modifying SC18 to form NY18, Asp225 → Gly, is that Gal-1 of LSTc can no longer interact with the HA RBS (Gln226, Lys222, and amino acid 225), which allows Gal-1 to move away from the protein surface, also affecting the interaction between the Neu5Ac residue of LSTc and the RBS. A further second modification forming the avian-adapted AV18, Asp190 → Glu, permits Gal-1 of LSTc to re-engage with the HA RBS, interacting with Gly225 and Lys222. While GlcNAc can no longer interact with the RBS in the LSTc−AV18 complex, it is this interaction that is a key decider for human adaption (Figure 5). The interaction between LSTa and AV18, avian-adapted HA, is principally via Neu5Ac and Gal-1 of the receptor. The nonreducing end $\alpha(2\rightarrow3)$ linkage between Neu5Ac and Gal-1 in LSTa permits Glu190 and Pro186 to interact with the sialyl group of Neu5Ac and O6-Gal-1 of LSTa.

The interaction between the glycan receptors and HA imposes conformational constraints upon the glycan, which is characterized in terms of the θ angle and the glycosidic torsional angles. We also correlated this analysis with biochemical HA−glycan binding specificity and affinity to ensure that this analysis is consistent with available crystal structure information. Importantly, using the θ angle as a parameter to characterize the overall shape of the glycan, our study demonstrates key differences in the form of LSTa and LSTc when they are free versus their bound states (LSTc−SC18, −NY18, and −AV18 and LSTa−AV18 and −NY18), which have not been possible to "capture" through X-ray crystal structures.

Previously, we have noted that human receptors in their unbound state sample a conformational space that resembles both a cone-like topology ($\theta \geq 110°$) and an umbrella-like topology ($\theta < 100°$), whereas avian receptors exclusively sample a cone-like topology.[28] By analyzing HA−glycan cocrystal structures, we postulated that glycans binding to "avian-adapted" HA RBS (such as AV18) would impose constraints on the human receptor (for example, LSTc) to preferentially sample a cone-like topology, whereas glycans binding to "human-adapted" HA (such as SC18) would constrain the glycan to preferentially sample an umbrella-like topology.[18] On the basis of the NMR analyses and MD simulations presented here, we demonstrate the former hypothesis to be correct. The θ angle of LSTc (Figure 6A) samples two distinct populations in the unbound state corresponding to umbrella-like and cone-like topologies. Upon binding to AV18 HA, LSTc predominantly samples a cone-like topology ($\theta \sim 110°$); this is due to GlcNAc of LSTc not being able to interact with the HA RBS. On the other hand, binding to the RBS of SC18 or NY18 imposes constraints on LSTc such that the glycan exclusively samples an umbrella-like topology.

Furthermore, compared to free LSTc, the bound glycan has restricted glycosidic torsional angles, with the torsional states observed for the LSTc−SC18 and LSTc−NY18 complexes being similar, whereas the mutation of Asp190 to Gly, seen in AV18, which abolishes the interaction between GlcNAc of LSTc and HA, produces unique glycosidic angles for ϕ_2 and ψ_2 and for ϕ_3 and ψ_3, which are distinct from those seen in the LSTc−SC18 and LSTc−NY18 complexes. This restriction in the glycosidic torsional angles is also observed in the LSTa−AV18 complex, even though the receptor is dynamic in the RBS, as indicated by the multiple states observed at the end of the MD simulation (Figures 7 and 8 of the Supporting Information), whereas in the LSTa−NY18 complex, where the nonreducing end Neu5Ac is the sole point of interaction with HA, linkages ϕ_2 and ψ_2, ϕ_3 and ψ_3, and ϕ_4 and ψ_4 all have glycosidic torsional angles that are similar to those seen in the free glycan (Figure 14 of the Supporting Information).

This stronger restriction correlates with a greater number of LSTc−SC18 HA contacts, which can be observed in the NMR STD measurements and higher measured binding affinity compared to those of the LSTc−NY18 HA complex. The

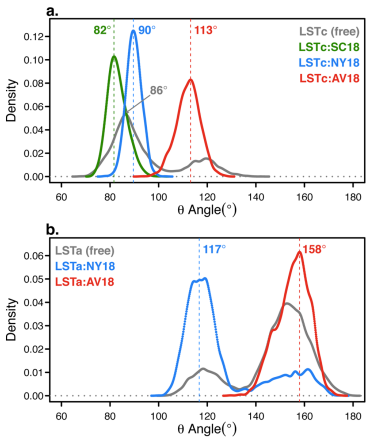


Figure 6. Density distribution plots of the topological θ angles of LSTc (A) and LSTa (B) free and bound to HA. The interaction of LSTc or LSTa with HA alters the topological θ angle assumed by the glycan receptor. The θ angles shown here are for the PCA-extracted conformers; a comparison of these with the θ angles for all of the conformers can be found in Figure 12 of the Supporting Information.

constraints on the avian receptor, LSTa, are consistent with the predominant contacts made by HA RBS with the Neu5Ac-α(2→3)-Gal motif as measured by NMR STD signals, which permit a higher degree of flexibility to the sugars on the reducing end of this terminal motif (as seen in the conformational map of ϕ_3 and ψ_3 and of ϕ_4 and ψ_4 in Figure 14 of the Supporting Information).

The results presented here clearly indicate that even one or two amino acid changes in the HA RBS impose different constraints on the conformation and topology of bound glycan receptors, which in turn governs the biochemical binding specificity and affinity. On the basis of this evidence, it is important to carefully assess the effects of transferring amino acid changes that lead to specific receptor binding properties for a given HA to, and from, a completely different strain or subtype. As stated earlier, the receptor specificity of avian- and human-adapted HAs has been broadly classified solely on the basis of a preference for α(2→3)- and α(2→6)-linked sialic acid. On the basis of the data presented here, we find that the amino acid composition of the RBS of various avian-adapted HA subtypes would critically govern structural constraints imposed on diverse and distinct sets of glycans expressed in different tissues, consistent with the ability of viruses from H5, H7, and H9 subtypes to infect distinct tissue types. On the other hand, human-adapted HAs share the characteristic binding to glycans expressed in human upper respiratory epithelia (particularly nonciliated goblet cells).[20,21,37] This characteristic binding can be explained on the basis of the structural constraints imposed by SC18 and NY18 on LSTc, which, in turn, is reflected by the θ parameter distribution. Given that goblet cells secrete mucins, it is possible that the characteristic binding of human-adapted HAs to these cell types would increase their propensity for aerosol formation and transmission.

Finally, the methods and framework presented in this study to measure the restriction imposed by the RBS of different HAs on the conformational space and topology sampled by glycan receptors can serve as a very useful tool for allowing more exact surveillance of emerging influenza viruses such as H7N9 and H5N1, to closely monitor their ability to bind to human receptors and acquire the capability for human-to-human transmission.

■ ASSOCIATED CONTENT

Supporting Information

Additional MD methods, supporting NMR spectra, and the multivariate analysis of the MD simulation data. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: rams@mit.edu.
*E-mail: guerrini@ronzoni.it.

Author Contributions

S.E., E.M., and T.R.R. made complementary and equal contributions to this work.

Funding

This work was funded in part by the National Institutes of Health (Grant R37 GM057073-13) and the National Research Foundation supported Interdisciplinary Research group in Infectious Diseases of SMART (Singapore MIT Alliance for Research and Technology). The 900 MHz spectra were recorded at the SONNMR Large Scale Facility in Utrecht, which was made possible by the financial support of the Access to Research Infrastructures activity in the seventh Framework Programme of the EC (Contract 261863, EU-NMR).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are very grateful for access to the NMR spectrometer (900 MHz) at the Utrecht NMR Facility and the assistance of Dr. Hans Wienk.

■ REFERENCES

(1) Ahmed, R., Oldstone, M. B., and Palese, P. (2007) Protective immunity and susceptibility to infectious diseases: Lessons from the 1918 influenza pandemic. *Nat. Immunol. 8*, 1188−1193.

(2) Perez Velasco, R., Praditsitthikorn, N., Wichmann, K., Mohara, A., Kotirum, S., Tantivess, S., Vallenas, C., Harmanci, H., and Teerawattananon, Y. (2012) Systematic review of economic evaluations of preparedness strategies and interventions against influenza pandemics. *PLoS One 7*, e30333.

(3) Fraser, C., Donnelly, C. A., Cauchemez, S., Hanage, W. P., Van Kerkhove, M. D., Hollingsworth, T. D., Griffin, J., Baggaley, R. F., Jenkins, H. E., Lyons, E. J., Jombart, T., Hinsley, W. R., Grassly, N. C., Balloux, F., Ghani, A. C., Ferguson, N. M., Rambaut, A., Pybus, O. G., Lopez-Gatell, H., Alpuche-Aranda, C. M., Chapela, I. B., Zavala, E. P., Guevara, D. M., Checchi, F., Garcia, E., Hugonnet, S., and Roth, C. (2009) Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science 324*, 1557−1561.

(4) Itoh, Y., Shinya, K., Kiso, M., Watanabe, T., Sakoda, Y., Hatta, M., Muramoto, Y., Tamura, D., Sakai-Tagawa, Y., Noda, T., Sakabe, S.,

Imai, M., Hatta, Y., Watanabe, S., Li, C., Yamada, S., Fujii, K., Murakami, S., Imai, H., Kakugawa, S., Ito, M., Takano, R., Iwatsuki-Horimoto, K., Shimojima, M., Horimoto, T., Goto, H., Takahashi, K., Makino, A., Ishigaki, H., Nakayama, M., Okamatsu, M., Warshauer, D., Shult, P. A., Saito, R., Suzuki, H., Furuta, Y., Yamashita, M., Mitamura, K., Nakano, K., Nakamura, M., Brockman-Schneider, R., Mitamura, H., Yamazaki, M., Sugaya, N., Suresh, M., Ozawa, M., Neumann, G., Gern, J., Kida, H., Ogasawara, K., and Kawaoka, Y. (2009) In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses. *Nature 460*, 1021−1025.

(5) Pearce, M. B., Jayaraman, A., Pappas, C., Belser, J. A., Zeng, H., Gustin, K. M., Maines, T. R., Sun, X., Raman, R., Cox, N. J., Sasisekharan, R., Katz, J. M., and Tumpey, T. M. (2012) Pathogenesis and transmission of swine origin A(H3N2)v influenza viruses in ferrets. *Proc. Natl. Acad. Sci. U.S.A. 109*, 3944−3949.

(6) Chen, Y., Liang, W., Yang, S., Wu, N., Gao, H., Sheng, J., Yao, H., Wo, J., Fang, Q., Cui, D., Li, Y., Yao, X., Zhang, Y., Wu, H., Zheng, S., Diao, H., Xia, S., Chan, K. H., Tsoi, H. W., Teng, J. L., Song, W., Wang, P., Lau, S. Y., Zheng, M., Chan, J. F., To, K. K., Chen, H., Li, L., and Yuen, K. Y. (2013) Human infections with the emerging avian influenza A H7N9 virus from wet market poultry: Clinical analysis and characterisation of viral genome. *Lancet 381*, 1916−1925.

(7) Fouchier, R. A., Schneeberger, P. M., Rozendaal, F. W., Broekman, J. M., Kemink, S. A., Munster, V., Kuiken, T., Rimmelzwaan, G. F., Schutten, M., Van Doornum, G. J., Koch, G., Bosman, A., Koopmans, M., and Osterhaus, A. D. (2004) Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *Proc. Natl. Acad. Sci. U.S.A. 101*, 1356−1361.

(8) Gao, R., Cao, B., Hu, Y., Feng, Z., Wang, D., Hu, W., Chen, J., Jie, Z., Qiu, H., Xu, K., Xu, X., Lu, H., Zhu, W., Gao, Z., Xiang, N., Shen, Y., He, Z., Gu, Y., Zhang, Z., Yang, Y., Zhao, X., Zhou, L., Li, X., Zou, S., Zhang, Y., Yang, L., Guo, J., Dong, J., Li, Q., Dong, L., Zhu, Y., Bai, T., Wang, S., Hao, P., Yang, W., Han, J., Yu, H., Li, D., Gao, G. F., Wu, G., Wang, Y., Yuan, Z., and Shu, Y. (2013) Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med. 368*, 1888−1897.

(9) Subbarao, K., and Katz, J. (2000) Avian influenza viruses infecting humans. *Cell. Mol. Life Sci. 57*, 1770−1784.

(10) Wan, H., Sorrell, E. M., Song, H., Hossain, M. J., Ramirez-Nieto, G., Monne, I., Stevens, J., Cattoli, G., Capua, I., Chen, L. M., Donis, R. O., Busch, J., Paulson, J. C., Brockwell, C., Webby, R., Blanco, J., Al-Natour, M. Q., and Perez, D. R. (2008) Replication and transmission of H9N2 influenza viruses in ferrets: Evaluation of pandemic potential. *PLoS One 3*, e2923.

(11) Watanabe, T., Kiso, M., Fukuyama, S., Nakajima, N., Imai, M., Yamada, S., Murakami, S., Yamayoshi, S., Iwatsuki-Horimoto, K., Sakoda, Y., Takashita, E., McBride, R., Noda, T., Hatta, M., Imai, H., Zhao, D., Kishida, N., Shirakura, M., de Vries, R. P., Shichinohe, S., Okamatsu, M., Tamura, T., Tomita, Y., Fujimoto, N., Goto, K., Katsura, H., Kawakami, E., Ishikawa, I., Watanabe, S., Ito, M., Sakai-Tagawa, Y., Sugita, Y., Uraki, R., Yamaji, R., Eisfeld, A. J., Zhong, G., Fan, S., Ping, J., Maher, E. A., Hanson, A., Uchida, Y., Saito, T., Ozawa, M., Neumann, G., Kida, H., Odagiri, T., Paulson, J. C., Hasegawa, H., Tashiro, M., and Kawaoka, Y. (2013) Characterization of H7N9 influenza A viruses isolated from humans. *Nature 501*, 551−555.

(12) Belser, J. A., Blixt, O., Chen, L. M., Pappas, C., Maines, T. R., Van Hoeven, N., Donis, R., Busch, J., McBride, R., Paulson, J. C., Katz, J. M., and Tumpey, T. M. (2008) Contemporary North American influenza H7 viruses possess human receptor specificity: Implications for virus transmissibility. *Proc. Natl. Acad. Sci. U.S.A. 105*, 7558−7563.

(13) Maines, T. R., Chen, L. M., Matsuoka, Y., Chen, H., Rowe, T., Ortin, J., Falcon, A., Nguyen, T. H., Mai le, Q., Sedyaningsih, E. R., Harun, S., Tumpey, T. M., Donis, R. O., Cox, N. J., Subbarao, K., and Katz, J. M. (2006) Lack of transmission of H5N1 avian-human reassortant influenza viruses in a ferret model. *Proc. Natl. Acad. Sci. U.S.A. 103*, 12121−12126.

(14) Gambaryan, A. S., Tuzikov, A. B., Piskarev, V. E., Yamnikova, S. S., Lvov, D. K., Robertson, J. S., Bovin, N. V., and Matrosovich, M. N. (1997) Specification of receptor-binding phenotypes of influenza virus isolates from different hosts using synthetic sialylglycopolymers: Non-egg-adapted human H1 and H3 influenza A and influenza B viruses share a common high binding affinity for 6′-sialyl(N-acetyllactosamine). *Virology 232*, 345−350.

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

(16) Shinya, K., Ebina, M., Yamada, S., Ono, M., Kasai, N., and Kawaoka, Y. (2006) Avian flu: Influenza virus receptors in the human airway. *Nature 440*, 435−436.

(17) Stevens, J., Blixt, O., Glaser, L., Taubenberger, J. K., Palese, P., Paulson, J. C., and Wilson, I. A. (2006) Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J. Mol. Biol. 355*, 1143−1155.

(18) Chandrasekaran, A., Srinivasan, A., Raman, R., Viswanathan, K., Raguram, S., Tumpey, T. M., Sasisekharan, V., and Sasisekharan, R. (2008) Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nat. Biotechnol. 26*, 107−113.

(19) Shriver, Z., Raman, R., Viswanathan, K., and Sasisekharan, R. (2009) Context-specific target definition in influenza A virus hemagglutinin-glycan receptor interactions. *Chem. Biol. 16*, 803−814.

(20) Srinivasan, A., Viswanathan, K., Raman, R., Chandrasekaran, A., Raguram, S., Tumpey, T. M., Sasisekharan, V., and Sasisekharan, R. (2008) Quantitative biochemical rationale for differences in transmissibility of 1918 pandemic influenza A viruses. *Proc. Natl. Acad. Sci. U.S.A. 105*, 2800−2805.

(21) Jayaraman, A., Chandrasekaran, A., Viswanathan, K., Raman, R., Fox, J. G., and Sasisekharan, R. (2012) Decoding the distribution of glycan receptors for human-adapted influenza A viruses in ferret respiratory tract. *PLoS One 7*, e27517.

(22) Bewley, C. A. (2008) Illuminating the switch in influenza viruses. *Nat. Biotechnol. 26*, 60−62.

(23) Tumpey, T. M., Maines, T. R., Van Hoeven, N., Glaser, L., Solorzano, A., Pappas, C., Cox, N. J., Swayne, D. E., Palese, P., Katz, J. M., and Garcia-Sastre, A. (2007) A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. *Science 315*, 655−659.

(24) Lakdawala, S. S., Shih, A. R., Jayaraman, A., Lamirande, E. W., Moore, I., Paskel, M., Sasisekharan, R., and Subbarao, K. (2013) Receptor specificity does not affect replication or virulence of the 2009 pandemic H1N1 influenza virus in mice and ferrets. *Virology 446*, 349−356.

(25) Roldos, V., Canada, F. J., and Jimenez-Barbero, J. (2011) Carbohydrate-protein interactions: A 3D view by NMR. *ChemBioChem 12*, 990−1005.

(26) Stevens, J., Blixt, O., Tumpey, T. M., Taubenberger, J. K., Paulson, J. C., and Wilson, I. A. (2006) Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science 312*, 404−410.

(27) Stevens, J., Corper, A. L., Basler, C. F., Taubenberger, J. K., Palese, P., and Wilson, I. A. (2004) Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. *Science 303*, 1866−1870.

(28) Sassaki, G. L., Elli, S., Rudd, T. R., Macchi, E., Yates, E. A., Naggi, A., Shriver, Z., Raman, R., Sasisekharan, R., Torri, G., and Guerrini, M. (2013) Human (α2→6) and avian (α2→3) sialylated receptors of influenza A virus show distinct conformations and dynamics in solution. *Biochemistry 52*, 7217−7230.

(29) Zhang, W., Shi, Y., Qi, J., Gao, F., Li, Q., Fan, Z., Yan, J., and Gao, G. F. (2013) Molecular basis of the receptor binding specificity switch of the hemagglutinins from both the 1918 and 2009 pandemic influenza A viruses by a D225G substitution. *J. Virol. 87*, 5949−5958.

(30) Gamblin, S. J., Haire, L. F., Russell, R. J., Stevens, D. J., Xiao, B., Ha, Y., Vasisht, N., Steinhauer, D. A., Daniels, R. S., Elliot, A., Wiley, D. C., and Skehel, J. J. (2004) The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science 303*, 1838−1842.

(31) Liu, J., Stevens, D. J., Haire, L. F., Walker, P. A., Coombs, P. J., Russell, R. J., Gamblin, S. J., and Skehel, J. J. (2009) Structures of receptor complexes formed by hemagglutinins from the Asian Influenza pandemic of 1957. *Proc. Natl. Acad. Sci. U.S.A. 106*, 17175−17180.
(32) Amadei, A., Linssen, A. B., and Berendsen, H. J. (1993) Essential dynamics of proteins. *Proteins 17*, 412−425.
(33) Kitao, A., and Go, N. (1999) Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol. 9*, 164−169.
(34) Gotsev, M. G., and Ivanov, P. M. (2009) Molecular dynamics of large-ring cyclodextrins: principal component analysis of the conformational interconversions. *J. Phys. Chem. B 113*, 5752−5759.
(35) Xu, D., Newhouse, E. I., Amaro, R. E., Pao, H. C., Cheng, L. S., Markwick, P. R., McCammon, J. A., Li, W. W., and Arzberger, P. W. (2009) Distinct glycan topology for avian and human sialopentasaccharide receptor analogues upon binding different hemagglutinins: A molecular dynamics perspective. *J. Mol. Biol. 387*, 465−491.
(36) Azzalini, A., Menardi, G., and Rosolin, T. (2012) *R package pdfCluster: Cluster analysis via nonparametric density estimation*, version 1.0-0, Universitá di Padova, Padua, Italy.
(37) Matrosovich, M. N., Matrosovich, T. Y., Gray, T., Roberts, N. A., and Klenk, H. D. (2004) Human and avian influenza viruses target different cell types in cultures of human airway epithelium. *Proc. Natl. Acad. Sci. U.S.A. 101*, 4620−4624.