# Avenues to Characterize the Interactions of Extended N-Glycans with Proteins by NMR Spectroscopy: The Influenza Hemagglutinin Case

*Beatriz Fernández de Toro[+], Wenjie Peng[+], Andrew J. Thompson, Gema Domínguez, F. Javier Cañada, Javier Pérez-Castells, James C. Paulson,\* Jesús Jiménez-Barbero,\* and Ángeles Canales\**

*Dedicated to Prof. Manuel Martín-Lomas on the occasion of his 77th birthday*

***Abstract:*** *Long-chain multiantenna N-glycans are extremely complex molecules. Their inherent flexibility and the presence of repetitions of monosaccharide units in similar chemical environments hamper their full characterization by X-ray diffraction or standard NMR methods. Herein, the successful conformational and interaction analysis of a sialylated tetradecasaccharide N-glycan presenting two LacNAc repetitions at each arm is presented. This glycan has been identified as the receptor of the hemagglutinin protein of pathogenic influenza viruses. To accomplish this study, a N-glycan conjugated with a lanthanide binding tag has been synthesized, enabling analysis of the system by paramagnetic NMR. Under paramagnetic conditions, the NMR signals of each sugar unit in the glycan have been determined. Furthermore, a detailed binding epitope of the tetradecasaccharide N-glycan in the presence of HK/68 hemagglutinin is described.*

Long-chain carbohydrates are common in nature, both in linear or branched structures. Today it is well-established that although glycans encode biological information in their structure, decoding this information is still a challenge. Recent studies point out that certain glycan motifs form secondary structure elements with conserved hydrogen bonds that resemble protein secondary structures.[1] However, there is a lack of knowledge about how these motifs are organized in space in the context of long-chain glycans. It is therefore challenging to predict the three-dimensional conformation of a glycan and its interactions with proteins. In this context, access to new methods both from the synthetic and structural viewpoints is essential to carry our systematic studies with complex glycans. Currently, such studies are a major task, since authentic N-glycan structures are challenging to synthesize, and their intrinsic chemical complexity precludes the use of the common structural chemistry techniques. N-glycans are highly flexible, hindering crystallization efforts.[2]

Although NMR spectroscopy has successfully been used to study linear oligosaccharides with up to five monosaccharide units,[3] for long-chain multiantennary N-glycans, the chemical equivalence of many of their NMR-active nuclei makes their conformational analysis and the exploration of their interaction properties rather challenging, especially with regard to branch specificity, molecular recognition features, and epitope characterization.

This is particularly magnified for N-glycans featuring multiple LacNAc (N-acetyllactosamine) repeats, since they contain pseudo-symmetric structures, with considerable overlap in regular NMR spectra. However, these molecules have key roles in nature. Sialylated long-chain glycans are recep-

[*] Dr. Á. Canales
Dpto Química Orgánica I, Fac. Ciencias Químicas, Universidad Complutense de Madrid
Avd. Complutense s/n, 28040 Madrid (Spain)
E-mail: ma.canales@quim.ucm.es

B. Fernández de Toro,[+] Prof. F. J. Cañada
Dpto Biología Estructural y Química
Centro de Investigaciones Biológicas, CIB-CSIC
C/Ramiro de Maeztu 9, 28040 Madrid (Spain)

Dr. W. Peng[+]
Key Laboratory of Systems Biomedicine (Ministry of Education),
Shanghai Center for Systems Biomedicine
Shanghai Jiao Tong University
800 Dongchuan Road, Shanghai 200240 (China)

Dr. W. Peng,[+] Dr. A. J. Thompson, Prof. J. C. Paulson
Depts. of Molecular Medicine, and Immunology and Microbiology
The Scripps Research Institute
10550 N. Torrey Pines Road, La Jolla, CA 92037 (USA)
E-mail: jpaulson@scripps.edu

Dr. G. Domínguez, Prof. J. Pérez-Castells
Dpto Química, Fac. Farmacia, Universidad San Pablo CEU
Madrid (Spain)

Prof. J. Jiménez-Barbero
CIC bioGUNE
Bizkaia Science and Technology Park, 48160 Bilbao (Spain)
and
IKERBASQUE, Basque Foundation for Science
48009 Bilbao (Spain)
E-mail: jjbarbero@cicbiogune.es

[+] These authors contributed equally to this work.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
https://doi.org/10.1002/anie.201807162.

tors for the influenza virus hemagglutinin, a pathogen that has evolved to efficiently transmit among individuals and between different species.

Seasonal flu (influenza virus infection) is a major threat to human health.[4] Influenza viruses, spread among individuals within airborne respiratory droplets, bind to sialic acid-containing glycan receptors on the surface of host epithelial cells. Receptor binding, the first step in the infection cycle, is mediated by the viral surface hemagglutinin (HA), while release of the virus from the infected cell is mediated by the neuraminidase (NA). Influenza viruses with different HAs (H1–18) and NAs (N1–11) circulate in avian species (birds), infect mammals (pigs, dogs, cats, bats) that come into contact with birds and humans, and occasionally infect and adapt to humans, causing worldwide pandemics. One barrier for transmission of avian viruses in humans is receptor specificity. HAs from human viruses recognize terminal Neu5Acα2-6Gal (human-type) receptor linkages abundant on epithelial cells, while avian HAs are specific for Neu5Acα2-3Gal (avian-type) receptors.[5] Although current human influenza viruses (H1N1 and H3N2) originated from avian viruses, only two amino acid mutations are required to switch the specificity of the H1 and H3 HAs from avian to human-type receptors.

Sialic acids, α2-3 or α2-6 linked to galactose (Gal), are found on a diverse array of glycoproteins and glycolipid glycans. While the widely used terms human-type and avian-type specificity are sufficient in many contexts, they over-simplify the underlying structural complexity of glycans at the cell surface. Glycomics profiling of the airway tissues or epithelial cells from human,[6] ferret,[7] and swine[8] has revealed unusual enrichment of large N-glycans with multiple LacNAc units, suggesting a potential role as viral receptors. In recent microarray and infection studies, it has been shown that contemporary H3N2 viruses have evolved specificity for α2-6-linked N-glycans with poly-LacNAc chains, while early viruses bound to human-type receptors regardless of length or branching. We reasoned that N-glycans with poly-LacNAc extensions are capable of forming bidentate binding inter-actions with two subunits of a single HA trimer.[9] This multivalent binding mode effectively increases avidity, con-tributing to transmission efficiency. To further investigate the role of LacNAc extensions during influenza virus infection, deep atomic-resolution analyses of the interactions between N-glycans and HA are needed.

In this context, we herein describe the first structural studies with a N-glycan containing poly-LacNAc repeats by using paramagnetic NMR. The use of glycans bearing LBT allows pseudo contact shifts (PCS) to be obtained; these are NMR parameters that depend on the distance and orientation between the paramagnetic lanthanide and the NMR-active nuclei and provide key structural data.[10] Since the different monosaccharide units along the two arms of the biantennary N-glycan display unique geometrical parameters (distance and orientation) with respect to the metal, different PCSs can be measured for the different monosaccharide moieties along the chains and at the different arms,[11] breaking their intrinsic chemical shift degeneracy. Protons closer to the LBT display larger PCS values and are differentiated in HSQC spectra. Thus, we rationally designed two truncated N-glycan sialo-
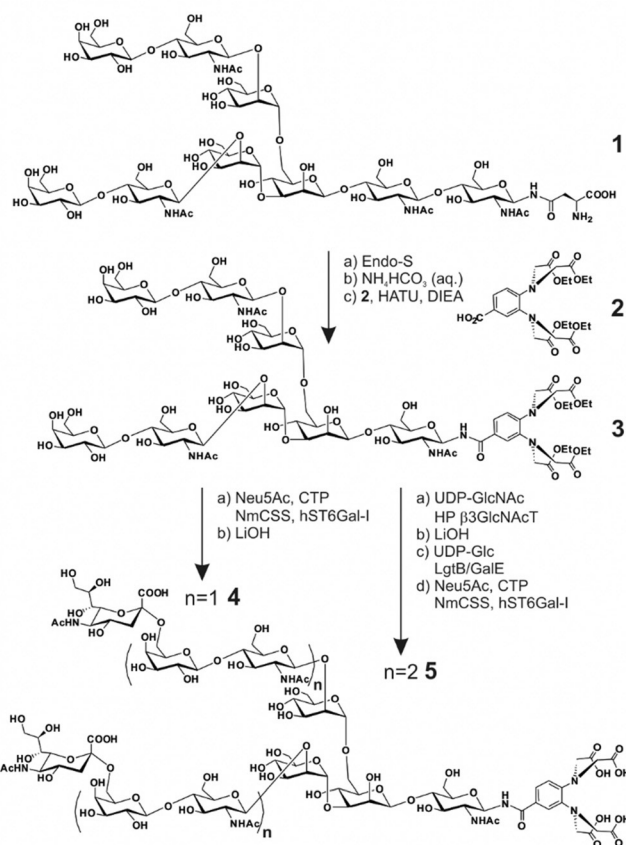


Scheme 1. Chemo-enzymatic synthesis of the tagged N-glycans 4 and 5.

sides 4 and 5, from which a GlcNAc-Asn fragment is missing at the reducing end (Scheme 1). Thus, compared to the so-called natural N-glycan, the LBT group is closer to the non-reducing ends. However, since the changes are located at the reducing end, these molecules should still maintain the same biological functions and present identical epitopes to the HAs as the normal N-glycans.

The synthesis of N-glycans 4 and 5 with one or two LacNAc units at each antenna is illustrated in Scheme 1. Initially, asialo bi-antennary N-glycan 1[12] was treated with endoglycosidase S (Endo-S), which hydrolyses the chitobiose core of the Asn-linked glycan,[13] to give the truncated N-glycan 1a in excellent yield (Supporting Information, Scheme S1A). Glycan 3 with an aglycone lanthanide-binding tag (LBT) at the reducing end was obtained in moderate yield by stirring hemiacetal 1a in saturated $NH_4HCO_3$ aqueous for one week,[14] followed by lyophilizing and condensing with benzoic acid 2[11b] in the presence of HATU and DIEA. With the key intermediate 3 in hand, we applied the chemo-enzymatic strategy to prepare the sialoside probes (Support-ing Information, Scheme S1B). Accordingly, by incubation with sialic acid, CMP-sialic acid synthetase (NmCSS) and α2,6-sialyltransferase (hST6Gal-I), galactoside 3 was sialy-lated in α2,6-linkage to form sialoside 3a in 92% yield.[9] The ethyl esters were removed by stirring glycan 3a and LiOH in water to afford the final product 4. Furthermore, by treating glycan 3 with UDP-GlcNAc and H. pylori β3-N-acetyl-

glucosaminyltransferase,[15] two N-acetyl-glucosaminyl residues were transferred to galactoses at the non-reducing end to form glycan **3b**, followed by saponification with LiOH to give **3c** in 86% yield for two steps. Di-LacNAc extended biantennary glycan **3d** was obtained in excellent yield by incubation of glycan **3c** with UDP-Glc and bacterial β4-galactosyltransferase/UDP-4-Gal-epimerase fusion protein (LgtB/GalE).[16] Finally, α2,6-linked sialoside **5** was afforded in 93% yield by applying similar reaction as for sialoside **3a**. LBT was synthesized following the protocol described previously.[11b]

For the NMR studies, two $^1$H-$^{13}$C HSQC NMR spectra of decasaccharide N-glycan **4** were acquired. In the blank experiment, the LBT was loaded with lanthanum (diamagnetic conditions), while dysprosium (paramagnetic conditions) was employed to generate the PCS. Fittingly, the presence of $Dy^{3+}$ at the LBT permitted differentiation of the signals of both branches of the N-glycan. In fact, the four sugars in each branch were distinguished. Strikingly, the comparison of the diamagnetic and paramagnetic $^1$H–$^{13}$C HSQC NMR spectra shows that the unique set of signals for every pair of Neu5Ac, Gal and GlcNAc units in the $La^{3+}$-containing sample is split into two distinct sets of cross-peaks in the presence of $Dy^{3+}$ (Supporting Information, Figure S2).

Once the method was validated for N-glycan **4**, the protocol was challenged with a sialylated tetradecasaccharide N-glycan (**5**), presenting two LacNAc repetitions at each arm. The analysis of the corresponding $^1$H–$^{13}$C HSQC NMR spectra of **5** complexed with $Dy^{3+}$ also permitted identification of different NMR signals for every sugar unit.

Remarkably, even the Neu5Ac external residues, located far away from the paramagnetic lanthanide can be distinguished. For instance, H7 of one Neu5Ac unit located at ca. 37 Å from $Dy^{3+}$. Moreover, the signals for all Gal and GlcNAc units could be distinguished (Figure 1). Interestingly, for N-glycans **4** and **5**, different PCSs behavior is observed for some proton signals. For instance, for **5**, H4 of residues Gal 5′ and 7′ displayed downfield shifting, while H4 of Gal 5 and 7 moved upfield (Figure 1).

The obtained experimental PCS (see the Supporting Information for details and values) for N-glycans **4** and **5** were then compared to those estimated, employing the MSpin software,[17] for different geometries of the N-glycans to deduce their actual solution conformational distributions. These geometries were calculated using MD, which predicted the presence of a conformational equilibrium around the Manα(1–6) Manβ with three conformers: extended conformers (Ψ 180°) with gg and gt rotamers (ω 60° and 180° respectively), and the folded gg conformer (Ψ 90° ω 60°) (Figure 2 A).

No satisfactory fits were found between experimental PCS and those estimated for any single conformation, strongly suggesting the presence of a conformational distribution. It is worth mentioning that determining a conformational distribution from average data (experimental PCS are the average of the values for all the conformations) is an ill-posed inverse problem that allows for infinite solutions.[18] In this work, we have chosen the approach that considers the ensemble with the minimum number of conformations that
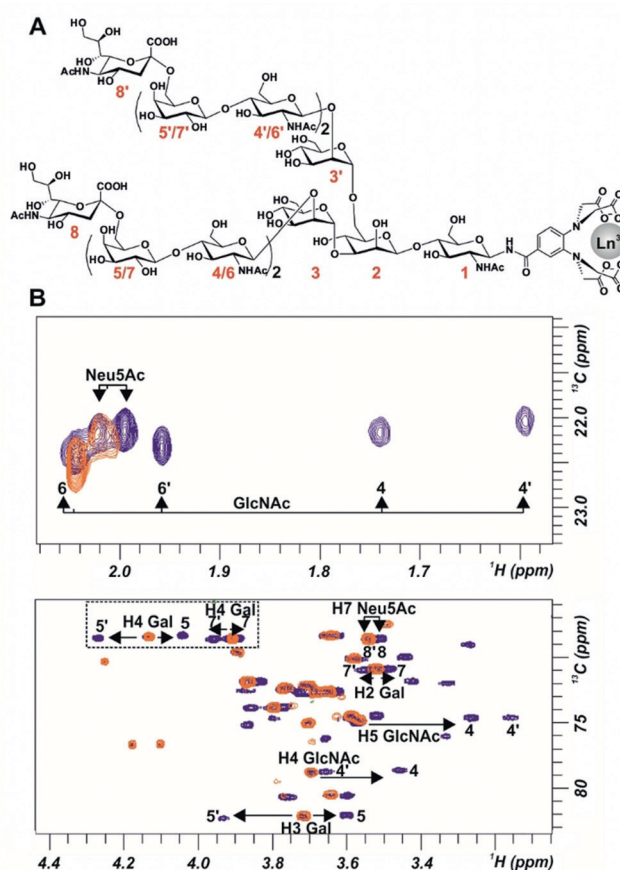


**Figure 1.** A) Structure of N-glycan **5** numbering each monosaccharide unit. B) Superimpositions of $^1$H–$^{13}$C HSQC spectra of N-glycan **5** loaded with lanthanum (diamagnetic, in orange) and with dysprosium (paramagnetic, in blue).

are able to explain the experimental data. A description of all the possible conformers according to MD simulations is given in the Supporting Information. In particular, for N-glycan **4**, the best fit was obtained for a two-state equilibrium around the Manα(1–6) Manβ linkage, including the extended conformers (Ψ 180°) with a major contribution (70%) of the gg rotamer (ω 60°) and a minor presence (30%) of the gt one (ω 180°). The associated quality factor (0.16) was excellent. In contrast, the best fit for N-glycan **5** strongly suggests the existence of a different conformational distribution around this angle, with three participating conformers equally populated: extended gg (Ψ 180° ω 60°), extended gt (Ψ 180° ω 180°) and folded gg (Ψ 90° ω 60°). The corresponding quality factor was also very good (0.24; Figure 2; Supporting Information, Tables S1, S2). According to these data, there is certain conformational restriction around Ψ for the N-glycan with shorter branches, while the longer N-glycan is more flexible. Obviously, there is also certain mobility around the Neu5Ac(2–6) Gal torsions. Nevertheless, the hairpin shape of the sialylated N-glycans is confirmed by the NOE between H3ax of the Neu5Ac unit and the GlcNAc Ac group of the vicinal LacNAc unit. This exclusive NOE permits to assess that the gt conformer is the major geometry adopted for Neu5Ac(2–6) Gal ω torsion, in agreement with the unre-
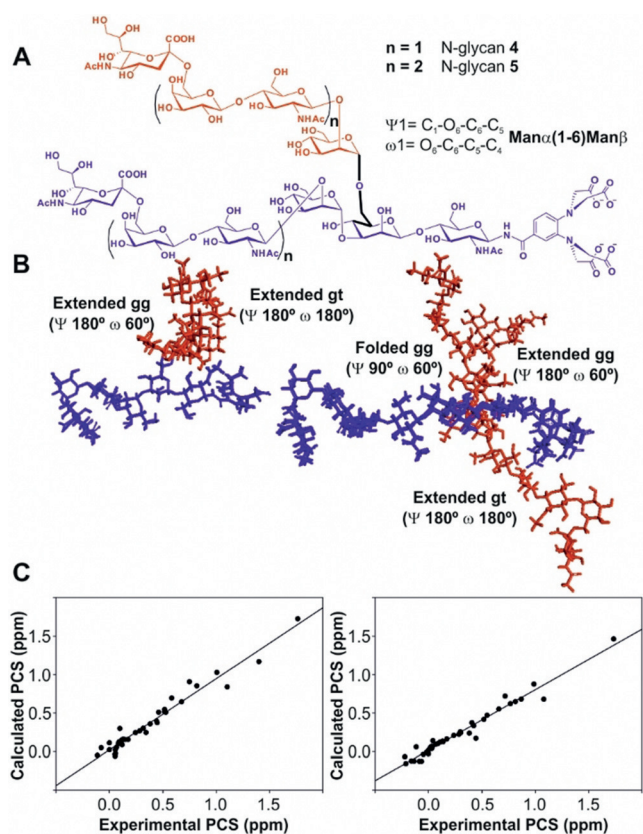
**Figure 2.** A) Structures of N-glycans **4** and **5** and definition of the torsion angles for the Manα (1–6) Manβ linkage. The B-arm is highlighted in red. B) Left: Superimposition of the extended gg and gt geometries of the Manα (1–6) Manβ linkage found in solution for N-glycan **4**, which provide the best fit to explain the experimental PCS. Right: Superimposition of the extended gg, extended gt, and folded gg geometries that provide the best fit for the PCS measured for N-glycan **5** (see individual structures in the Supporting Information, Figure S3). C) Correlation between the experimental PCS and the calculated values for a combination of the two conformations (gg:gt 70:30) for the N-glycan **4** (left) and for the three equally populated conformations of N-glycan **5** (right), as shown in panel B.



**Figure 3.** STD NMR experiment of N-glycan **5** under diamagnetic conditions (left) and loaded with Dy$^{3+}$ (right). The STD spectra (blue) show the signals of the protons that are involved in the interaction process. The H7 NMR signals of both Neu5Ac units 8 and 8′ are highlighted.

strained MD simulations (Supporting Information, Table S4). Interestingly, the NMR experimental data show that compound **4** is less flexible than the analogous asialobiantennary N-glycan previously described using this method.[11b]

Thus, the longer chains can efficiently explore a wider conformational space. The higher inherent flexibility of the carbohydrate chain may be a key factor to explain the preferential recognition of longer glycans by HA viral proteins.[9,16]

Next, the interaction of the di-LacNAc-containing sialoside **5** with a recombinant H3 hemagglutinin from A/Hong Kong/1/1968 (HK/68) H3N2 influenza virus was evaluated. Based on previous avidity assays for HK/68,[9] the N-glycan with one LacNAc (similar to **4**) displays weak binding, while a glycan with two LacNAc repeats like N-glycan **5** binds with higher avidity. From the NMR perspective, STD data obtained of N-glycan **5** under diamagnetic conditions also showed glycan binding to the hemagglutinin, involving the Neu5Ac residues, but no discrimination of the specific
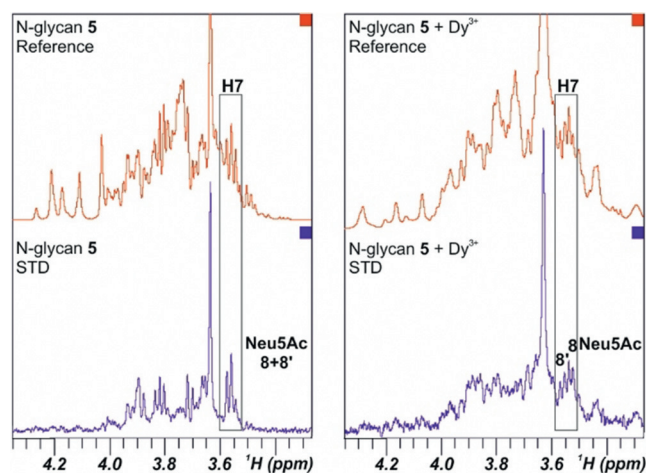
involvement of A and/or B branches could be evidenced (Figure 3 left) owing to the chemical-shift degeneracy. However, the dispersion of the signals of **5** in the 1D $^1$H NMR spectrum in the presence of the paramagnetic ion (Dy$^{3+}$) was excellent. Therefore, STD experiments were employed to unravel specific molecular recognition features for the two individual branches. Indeed, STDs were observed for the different $^1$H NMR signals for each Neu5Ac moieties (Figure 3 right). Thus, the interaction of both Neu5Ac residues, 8 and 8′, with HK/68 was demonstrated in a non-ambiguous manner. Obviously, the obtained data indicate that both arms effectively interact with the hemagglutinin, opening the possibilities to the establishment of cluster effects.

We[9,16] and others[19] have reported that some human influenza viruses preferentially recognize longer α2,6-linked N-glycan sialosides. However, the detailed interactions between the HAs and the large sialoside receptors have not yet been closely investigated. Herein, we have designed and chemoenzymatically synthesized two biantennary N-glycan sialosides **4** and **5**, containing one or two LacNAc repeat units, respectively.

By applying the novel paramagnetic NMR strategy for glycans, we have obtained PCS values for every sugar unit for both sialosides, even for the terminal Neu5Ac residues. Subsequent data analysis has revealed that sialoside **5** is more flexible than **4**. The synergic combination of paramagnetic NMR with STD NMR studies has demonstrated, in a non-ambiguous manner, that both sialic acids of **5** can interact independently with H3N2 HK/68.

In conclusion, we have achieved the NMR characterization of long-chain glycans up to 46 Å in length, including the resolution of NMR signals of individual monosaccharide units in biantennary N-glycans with two LacNAc repeats at each arm. The extension to more complex glycans represents a further challenge that we are currently addressing, including the synthesis of N-glycans with tri-LacNAc repeats decorated with LBTs. We anticipate that this novel NMR method will be

© 2018 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

a significant breakthrough, enabling analysis of simultaneous/individual binding of very complex glycans to influenza hemagglutinins. This opens new avenues to deduce epitope selection at the atomic level and eventually to understand cluster effects and how influenza viruses recognize, bind, and infect host cells.

## Conflict of interest

The authors declare no conflict of interest.

[1] T. Aeschbacher, M. Zierke, M. Smieško, M. Collot, J. M. Mallet, B. Ernst, H. T. Allain Frédéric, M. Schubert, *Chem. Eur. J* **2017**, *23*, 11598–11610.

[2] J. Agirre, J. Iglesias-Fernández, C. Rovira, G. J. Davies, K. S. Wilson, K. D. Cowtan, *Nat. Struct. Mol. Biol.* **2015**, *22*, 833.

[3] E. Macchi, T. R. Rudd, R. Raman, R. Sasisekharan, E. A. Yates, A. Naggi, M. Guerrini, S. Elli, *Biochemistry* **2016**, *55*, 6605–6616.

[4] a) https://www.cdc.gov/flu/; b) V. N. Petrova, C. A. Russell, *Nat. Rev. Microbiol.* **2017**, *16*, 47–60.

[5] G. N. Rogers, J. C. Paulson, *Virology* **1983**, *127*, 361–373.

[6] T. Walther, R. Karamanska, R. W. Y. Chan, M. C. W. Chan, N. Jia, G. Air, C. Hopton, M. P. Wong, A. Dell, J. S. Malik Peiris, S. M. Haslam, J. M. Nicholls, *PLOS Pathog.* **2013**, *9*, e1003223.

[7] N. Jia, W. S. Barclay, K. Roberts, H.-L. Yen, R. W. Y. Chan, A. K. Y. Lam, G. Air, J. S. M. Peiris, A. Dell, J. M. Nicholls, S. M. Haslam, *J. Biol. Chem.* **2014**, *289*, 28489–28504.

[8] A. C. Bateman, R. Karamanska, M. G. Busch, A. Dell, C. W. Olsen, S. M. Haslam, *J. Biol. Chem.* **2010**, *285*, 34016–34026.

[9] W. Peng, R. P. de Vries, O. C. Grant, A. J. Thompson, R. McBride, B. Tsogtbaatar, P. S. Lee, N. Razi, I. A. Wilson, R. J. Woods, J. C. Paulson, *Cell Host Microbe* **2017**, *21*, 23–34.

[10] G. Otting, *J. Biomol. NMR* **2008**, *42*, 1–9.

[11] a) A. Canales, I. Boos, L. Perkams, L. Karst, T. Luber, T. Karagiannis, G. Domínguez, F. J. Cañada, J. Pérez-Castells, D. Häussinger, C. Unverzagt, J. Jiménez-Barbero, *Angew. Chem. Int. Ed.* **2017**, *56*, 14987–14991; *Angew. Chem.* **2017**, *129*, 15183–15187; b) A. Canales, A. Mallagaray, J. Pérez-Castells, I. Boos, C. Unverzagt, S. André, H.-J. Gabius, F. J. Cañada, J. Jiménez-Barbero, *Angew. Chem. Int. Ed.* **2013**, *52*, 13789–13793; *Angew. Chem.* **2013**, *125*, 14034–14038.

[12] W. Peng, J. C. Paulson, *J. Am. Chem. Soc.* **2017**, *139*, 12450–12458.

[13] M. Collin, A. Olsén, *EMBO J.* **2001**, *20*, 3046.

[14] R. Chen, T. J. Tolbert, *J. Am. Chem. Soc.* **2010**, *132*, 3211–3216.

[15] W. Peng, J. Pranskevich, C. Nycholat, M. Gilbert, W. Wakarchuk, J. C. Paulson, N. Razi, *Glycobiology* **2012**, *22*, 1453–1464.

[16] C. M. Nycholat, R. McBride, D. C. Ekiert, R. Xu, J. Rangarajan, W. Peng, N. Razi, M. Gilbert, W. Wakarchuk, I. A. Wilson, J. C. Paulson, *Angew. Chem. Int. Ed.* **2012**, *51*, 4860–4863; *Angew. Chem.* **2012**, *124*, 4944–4947.

[17] a) A. Navarro-Vázquez, *Magn. Reson. Chem.* **2012**, *50*, S73–S79; b) https://www.mestrelab.com.

[18] a) E. Ravera, L. Sgheri, G. Parigi, C. Luchinat, *Phys. Chem. Chem. Phys.* **2016**, *18*, 5686–5701; b) M. Bonomi, G. T. Heller, C. Camilloni, M. Vendruscolo, *Curr. Opin. Struct. Biol.* **2017**, *42*, 106–116.

[19] H. Yang, P. J. Carney, J. C. Chang, Z. Guo, J. M. Villanueva, J. Stevens, *Virology* **2015**, *477*, 18–31.