

# Glycosylation and the global virome

Cassandra L. Pegg<sup>1</sup>  | Benjamin L. Schulz<sup>1</sup>  | Benjamin A. Neely<sup>2</sup>  |  
Gregory F. Albery<sup>3</sup> | Colin J. Carlson<sup>3,4,5</sup> 

<sup>1</sup>School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Queensland, Australia

<sup>2</sup>National Institute of Standards and Technology, Charleston, South Carolina, USA

<sup>3</sup>Department of Biology, Georgetown University, Washington, District of Columbia, USA

<sup>4</sup>Department of Microbiology and Immunology, Georgetown University Medical Center, Washington, District of Columbia, USA

<sup>5</sup>Center for Global Health Science and Security, Georgetown University Medical Center, Washington, District of Columbia, USA

## Correspondence

Cassandra L. Pegg, School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, QLD 4072, Australia.

Email: [cassandra.pegg@uqconnect.edu.au](mailto:cassandra.pegg@uqconnect.edu.au)

Colin J. Carlson, Department of Biology, Georgetown University, Washington, DC 20007, USA.

Email: [colin.carlson@georgetown.edu](mailto:colin.carlson@georgetown.edu)

## Funding information

National Health and Medical Research Council, Grant/Award Number: APP1186699; National Science Foundation, Grant/Award Number: BIL 2021909

**Handling Editor:** Camille Bonneaud

## Abstract

The sugars that coat the outsides of viruses and host cells are key to successful disease transmission, but they remain understudied compared to other molecular features. Understanding the comparative zoology of glycosylation - and harnessing it for predictive science - could help close the molecular gap in zoonotic risk assessment.

## KEYWORDS

glycosylation, Microbial Biology, Modelling, Molecular Evolution, Viral Ecology, Viral Evolution

## 1 | INTRODUCTION

Due to recent encounters with zoonotic viruses like Ebola virus and SARS-CoV-2, efforts to forecast the zoonotic risk of wild-life viruses - and, more broadly, to understand the biological constraints on cross-species transmission - are increasingly appealing (Albery et al., 2021). To date, most of these efforts rely on easily-observed traits of hosts, like morphology, diet, or phylogeny (Albery et al., 2020; Han et al., 2016). Despite their distance from the molecular determinants of transmission, these traits can be used to build models that have surprising predictive accuracy (Becker et al., 2022).

Many microbiologists have expressed a healthy scepticism of these approaches, which often entirely lack predictors that consider the molecular biology of hosts or any viral traits, and therefore, only coarsely infer the cellular processes of infection through other proxies that are correlated across evolutionary space. As a rare exception, genomic approaches are increasingly being used to close this gap (Babayan et al., 2018; Mollentze et al., 2021), and can help identify salient mechanisms of host-virus interactions (e.g., CpG dinucleotide depletion in vertebrate viruses appears to help them evade innate immune responses like the zinc finger antiviral protein [Takata et al., 2017]). However, as predictive features, genomic traits are often confounded by evolutionary signals (Di Giallonardo

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

et al., 2017; Shackelton et al., 2006), and genomic data only offer limited insights into the actual three-dimensional structural compatibility of viral and host cell surfaces, a “lock-and-key” type process. This lock-and-key interaction not only allows efficient infection of susceptible hosts, but also limits cross-species viral transmission. Structural modelling approaches have been used to examine the binding of viral proteins and host cell receptors, most recently in the context of research on SARS-CoV-2, but many of these simulations neglect key information: the glycosylation of these structures. Indeed, the handful that do address this aspect have revealed unexpected and important roles for glycosylation in these processes including those for SARS-CoV-2 (Casalino et al., 2020; Ghorbani et al., 2021; Sztain et al., 2021; Zhao et al., 2020), HIV (Berndsen et al., 2020; Ferreira et al., 2018; Lemmin et al., 2017; Stewart-Jones et al., 2016; Wood et al., 2013; Yang et al., 2017), and influenza viruses (Kasson & Pande, 2008; Newhouse et al., 2009; Seitz et al., 2020; Xu et al., 2009).

The sugars, or glycans, that decorate host cell surface macromolecules are often critical ligands that viruses associate with to enter cells (Jones et al., 2021). Viral proteins from different virus families exhibit substantial variation in binding affinity towards host glycan receptors, and the compatibility between different viral proteins and host glycosylation varies between host and virus species, across tissues and organ systems, and even over time or between individuals (Jones et al., 2021; Maginnis, 2018; Thompson et al., 2019). Some viral surface proteins become glycosylated by host cell machinery during infection, and in the process can mimic host cell surfaces or shield proteins from antibody recognition, helping viruses evade the host's immune system (Bagdonaite et al., 2018; Bagdonaite & Wandall, 2018; Watanabe et al., 2019; Zhao et al., 2021). The glycans on viral surfaces are also recognized by host glycan-binding proteins on immune cells that capture viruses, either preventing or promoting infection (Crocker et al., 2007; Erikson et al., 2015). These aspects of host-virus compatibility can create or unlock barriers to transmission, but are poorly characterized as an underlying structural determinant of host-virus networks because glycan structures are subject to rapid regulation and are sometimes perceived as being analytically challenging. We suggest an undertaking to describe the comparative zoology of glycoproteins, and their role in structuring the global virome.

## 2 | WHAT SUGARS DO, AND HOW

Glycans are a key feature of the cell surface and extracellular matrix of eukarya, archaea, and bacteria (West et al., 2021). Glycans can be found on proteins or lipids, and technological advances (Everest-Dass et al., 2018) in the 21st century have greatly increased our ability to identify, characterize, and manipulate glycosylation, which has in turn supported deeper insights into the multitude of diverse roles it plays (Moremen et al., 2012). Glycans can constitute a substantial proportion of the molecular mass of a protein (Varki, 2017), contributing to their biophysical properties and influencing protein targeting, folding, structure and secretion (Ohtsubo & Marth, 2006;

Varki, 2017). Moreover, glycans are mostly located on secreted proteins and at the cell surface, and are therefore key determinants in molecular recognition events (Schjoldager et al., 2020).

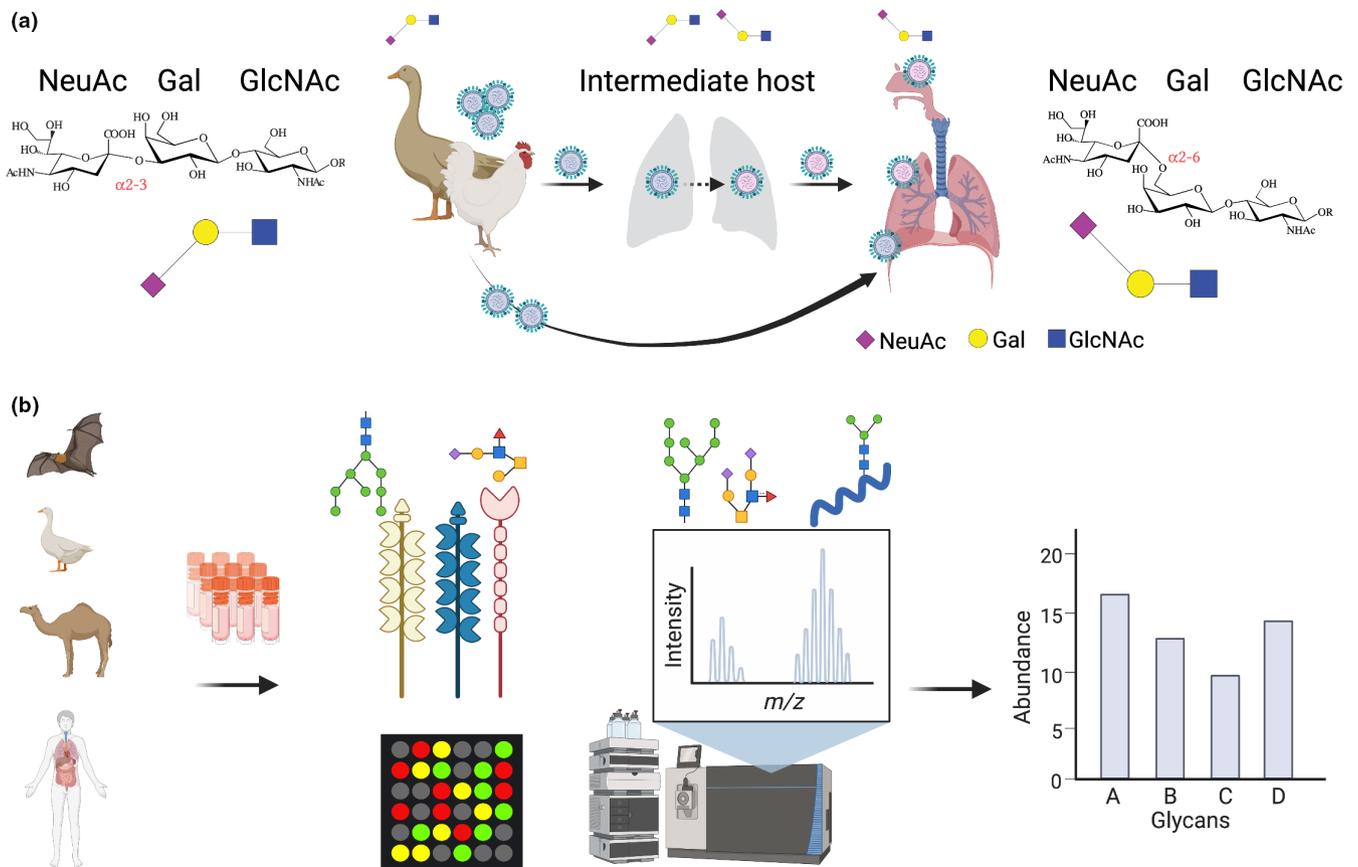
Glycans exhibit tremendous structural diversity (Spiro, 2002). Unlike proteins, whose sequences can be predicted from gene sequences, the biosynthesis of glycans is not directly template-driven; glycans are built, modified and trimmed by an extensive network of co-expressed enzymes that are differentially expressed in cells and tissues, and can be affected by factors intrinsic and extrinsic to the cell. Glycan structures are therefore specific to various organisms, tissues, and cells, and the resulting structures can be highly heterogeneous, imparting additional complexity to the structural and functional properties of proteins (Schjoldager et al., 2020).

Glycans act as receptors, coreceptors or attachment factors for numerous viruses (Thompson et al., 2019) including HIV, dengue (Raman et al., 2016), MERS-CoV (Park et al., 2019), influenza (Shinya et al., 2006), and SARS-CoV-2 (Yang et al., 2020). For example, both avian and human-adapted influenza viruses bind to glycans that terminate with sialic acid, but avian influenza preferentially binds to sialic acid with  $\alpha$ 2-3-linkages, while human-adapted influenza prefers  $\alpha$ 2-6-linkages, which are expressed in the human upper respiratory tract (Kuchipudi et al., 2009; Shinya et al., 2006). When avian-origin influenza lineages jump directly into humans, infections mostly become established in the lower lungs where cells express  $\alpha$ 2-3-linkages, leading to rare infections that are generally more severe but less transmissible (Stevens et al., 2006). Animal bridge hosts that can be infected by different influenza subtypes (including avian species) or express both types of sialic acid - for example, swine - provide an environment where viral lineages may directly undergo mutation to adapt to  $\alpha$ 2-6-binding (Thompson & Paulson, 2021), or where human and avian lineages cocirculate and undergo reassortment (Figure 1a). These hosts therefore provide an evolutionary stepping stone for avian lineages to switch to  $\alpha$ 2-6-binding, opening a transmission route for harmful zoonoses and producing more transmissible strains that pose epidemic or pandemic threats in humans (Chothe et al., 2017).

In cases like these, glycosylation is a key driver of host range and zoonotic risk, but one that is often neglected or folded into the “black box” of host-virus interactions and evolutionary dynamics in systems that are less well-characterized than influenza. However, the general importance of glycosylation in host-pathogen interactions is well established (Suenaga & Arase, 2015; Watanabe et al., 2019). Given that potential differences in glycosylation presence and structure can have profound effects on molecular interactions, we suggest there is a clear need to measure the glycomes of potential hosts as part of broader efforts to describe viral ecology and emergence.

## 3 | CHARACTERIZING GLYCANS AT DIFFERENT SCALES

The inherent structural complexity and heterogeneity of glycans across species, individuals, organ systems, tissues, and even time and



**FIGURE 1** Glycosylation underlies the evolutionary shift towards zoonotic emergence of influenza. (a) When avian influenza makes the jump directly to humans, preferential binding to  $\alpha$ 2-3-linked sialic acid expressed in the lower lung leads to rarer infections that can be more severe but less transmissible. In intermediate hosts that express both  $\alpha$ 2-3-linked and  $\alpha$ 2-6-linked sialic acid, viruses can undergo an evolutionary shift that facilitates emergence in humans, who express  $\alpha$ 2-6-linked sialic acid in the upper respiratory tract. (b) An array of different tools can be used to study the glycosylation profiles of tissues, cells or proteins from a global level down to the glycan level. Figure created with BioRender ([biorender.com](https://biorender.com))

space make them an analytically challenging subject. Additionally, and in contrast to other biomolecules, they can have extraordinarily high structural complexity due to the variety of monosaccharide building blocks and the multiple ways they can attach to each other, both in bond configuration ( $\alpha$  and  $\beta$ ) and in the positions of the intersaccharide linkages within the molecules (Cummings, 2009). Nevertheless, a number of techniques can be used individually or in combination to characterize protein glycosylation, with mass spectrometry being a powerful and widely used tool that can be incorporated at various levels (Figure 1b). Generally, the overarching aim of glycosylation analyses is to deduce one or more of the following: the monosaccharide composition of the glycans, the order and branching of monosaccharides in a glycan, the types of glycosidic linkages and monosaccharide anomericity, or the location of the glycosylation sites on a protein. The functional roles of glycans can also be assessed by measuring noncovalent interactions between specific proteins and glycans. A global view can be obtained from antibody or lectin binding to selectively identify glycan epitopes or motifs in tissues, cells and proteins. Lectins can be used in array-based platforms (Gao et al., 2019) enabling high-throughput analyses with the caveat that they do not provide comprehensive structural information. The

most effective way to obtain structural details of glycans is to use a glycomic workflow whereby the glycans are chemically or enzymatically released from glycoproteins. The released glycans are typically chemically labelled with fluorescent tags and analysed by liquid chromatography or capillary electrophoresis with a mass spectrometer used for enhanced detection (Everest-Dass et al., 2018). These methods for studying the glycome can provide monosaccharide composition and sequence information, and at times, linkage position. Coupling these analytical techniques with enzymes that cleave specific monosaccharide linkages provides additional precise structural information.

A limitation of glycomic approaches is that protein- and site-specificity is lost with glycan release. Nevertheless, these techniques are the most powerful for providing the basic information about glycan structure that will probably form the basis of future predictive models (see below). This is because the same glycan structures or epitopes can often be found on many different sites and proteins from the same cell, because they share the same glycan biosynthetic pathways. In addition, a key benefit of glycomics, compared to proteomics or glycoproteomics (the study of proteins and glycosylated proteins, respectively), is that analysis of released

glycans does not typically require prior knowledge of the genome. Although glycan biosynthesis is nontemplate driven and millions of possible glycan structures can be predicted (Cummings, 2009), there are comparatively few glycan structures actually observed on glycoproteins (Werz et al., 2007), and these can be predicted or de novo structurally determined without knowledge of the genome (Kellman & Lewis, 2021). If the annotated genome is available, genomic and phylogenetic profiling of glycan metabolism enzymes can greatly assist or validate glycomic approaches (McVeigh et al., 2018), especially where the glycan profiles from species contain structures that have not yet been defined (West et al., 2021). Furthermore, mass spectrometry glycoproteomics can also be used to identify and measure peptides with attached glycans. In this case, fine structural detail of the glycan structure is lost but the site of attachment and the level of site occupancy is retained.

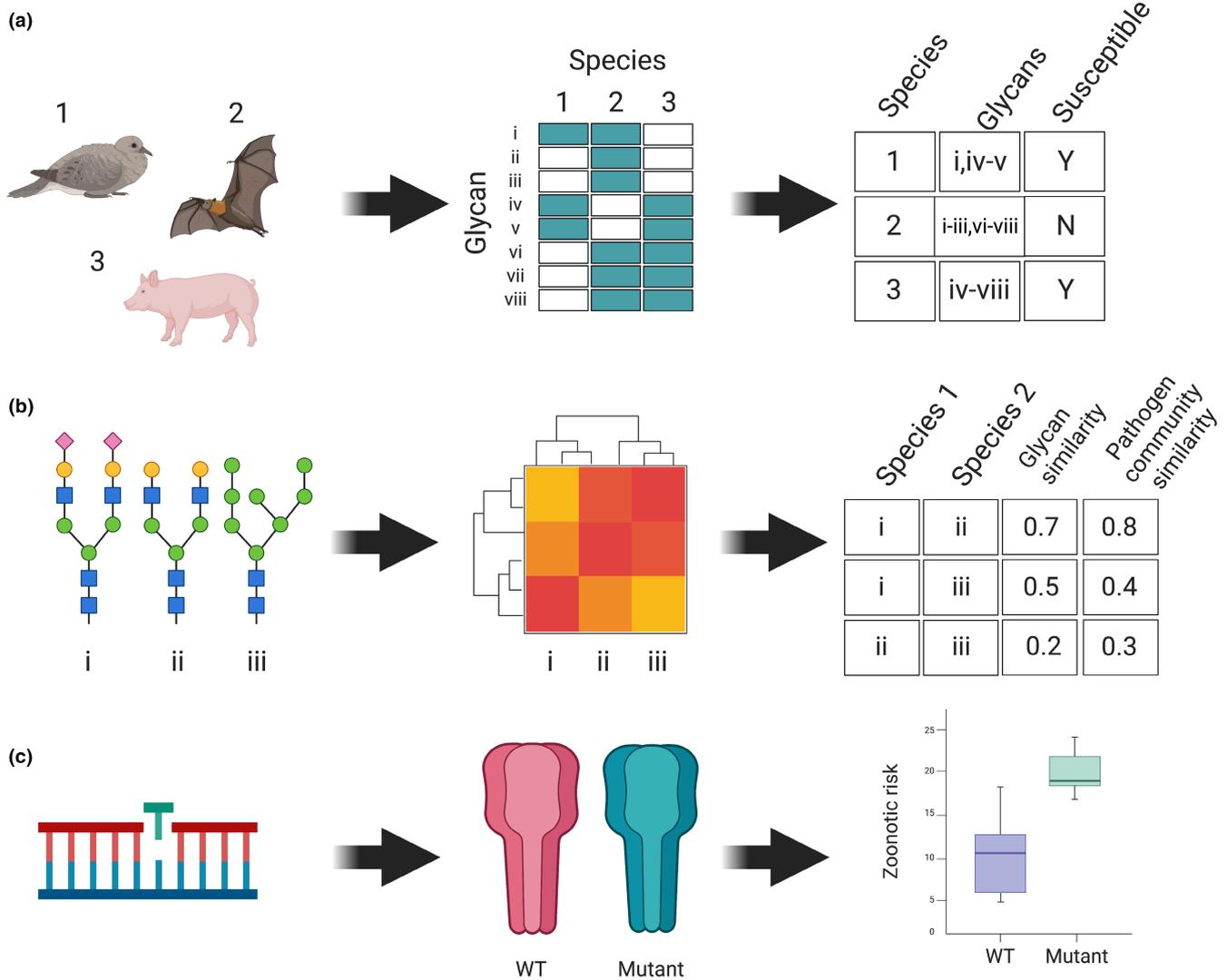
#### 4 | BUILDING GLYCOSYLATION INTO PREDICTIVE SCIENCE

Understanding the landscape of host glycosylation might help scientists build better predictive tools to understand the broader rules of viral cross-species transmission or even the special case of zoonotic risk. This could be accomplished in a number of ways, most of which are untested. Glycosylation could be represented as data in several ways, ranging from simple (e.g., the presence or absence of a specific set of glycan structures) to complex (generating quantitative features using graph representations of the glycan structure [Alonso et al., 2018]). These data can then be used several ways. For example, recent attention on SARS-CoV-2's use of the ACE2 receptor has sparked the development of models that predict host susceptibility based on receptor sequences (Fischhoff et al., 2021), but glycosylation is a missing element; incorporating glycan structures as receptor or coreceptor "metadata" might help researchers better understand viral attachment (Figure 2a). For instance, sialic acid and heparan sulphate are key cell surface glycans that are coreceptors for SARS-CoV-2 (Clausen et al., 2020; Nguyen et al., 2021). Conversely, from the host perspective, specific glycans may promote viral entry as is the case for SARS-CoV-2, where site-specific glycans of ACE2 have been implicated in receptor-viral binding (Zhao et al., 2020). Thus similar glycosylation might help explain pathogen sharing between different animals (Figure 2b), and (according to some preliminary evidence [Burkholz et al., 2021]) might even help unpack some of the microbiology inside the black box "phylogenetic distance effect" that broadly structures the viral sharing network (Albery et al., 2020). If viral glycosylation helps evade host immune system detection, these dimensions of similarity may help explain how particular bridge host "stepping stones" are possible, including in cases of zoonotic emergence (during early characterization, glycan motifs or structures can even be proactively searched to identify homology in human and animal hosts, and linked to viruses where glycan-binding preferences have been established). Similarly, understanding the role of viral mutations that alter the viral surface glycoproteins (Figure 2c) may

lead to more targeted insights about how glycosylation relates to zoonotic risk (as in the example of influenza and sialic acid; Figure 1a).

Most of these examples are still hypothetical, but in limited cases, these types of model-based exploration have shown tremendous promise. For example, a recent study used graph representations of glycans and multiple kinds of advanced machine learning (graph convolutional neural networks and natural language models) to predict host identity and glycan immunogenicity, and was able to predict influenza and rotavirus binding affinity for host receptors from different species (Burkholz et al., 2021). Studies like these are exciting proofs-of-concept and point to the idea that feature representations of glycans may eventually be useful as part of a broader palette of cell- and virus-level trait predictors used to make even more advanced (and crucially, microbiology-driven) machine learning or network models.

In order to power these kinds of approaches, more data is needed about the "global glycome." While the human glycome is well studied (Jia et al., 2020), the glycomes of animal reservoirs are severely understudied - a major problem when, for example, host range is extensive (e.g., influenza A virus can infect captive and wild animals including birds, dogs, cats, pigs, horses, bats, seals and even some reptiles [Short et al., 2015]). Where existing glycomic data sets are unavailable, transcriptomics coupled with experimentally-defined and predicted glycan biosynthetic pathways could help fill the gaps (Dworkin et al., 2022; Kellman & Lewis, 2021) while existing glycomic data sets can be used to train models. When expanding existing data sets, wild reservoir populations should be a priority for experimental glycosylation analyses, particularly those that are endemically infected, at the human-animal interface and those likely to act as an intermediate "mixing vessel" for cross-species transmission. Tissues, cells and fluids that are the routes of entry for infectious agents or are predicted to be involved in tropism and systemic spread should be a focal point of analysis. These may include the mucosal epithelial tissue of the gastrointestinal, urogenital and respiratory tracts and cells of the skin, lymphatic system and blood vessels. The types of data to be incorporated from these sample types could range from global glycome analyses to precise glycan structure and site occupancy information. Ultimately, the types of experiments conducted will be determined by the capabilities of the laboratory. In depth protocols are available for global structural (Jensen et al., 2012; Li et al., 2020) and protein- and site-specific (Hart & Wells, 2021; Kolarich et al., 2012; Oliveira et al., 2021; Remorozo et al., 2021; Riley et al., 2020) characterization of glycans. Given the versatility of liquid chromatography mass spectrometry to study glycosylation, this approach seems most accessible with costs predicted to be in the low to medium range depending on the level of structural characterization achieved. With unlimited sample availability and technical resources, multiple tissues from multiple species could be characterized and this data used for modelling. Tissue types may be limited due to the requirements for lethal or nonlethal sampling, the ability to dissect out specific tissues, or even the capacity to sample all species, though many field researchers have suitable samples already collected and in storage. As with most fieldwork, sample



**FIGURE 2** Adding glycosylation to microbiology-smart modelling. (a) Different animals have different combinations of glycans, which may help unpack how specific glycans contribute to susceptibility to a given pathogen. In this example, species 2's lack of glycans iv and v could explain its lack of susceptibility to the pathogen. (b) When the mechanism is understood in better detail, the glycosylation of a single structure (e.g., the ACE2 receptor) might help predict cross-species transmission potential for a specific virus (e.g., SARS-CoV-2) - if the structural similarity of a given glycan can be converted into machine-readable features. (c) Understanding mechanisms in greater detail may improve other kinds of predictions about cross-species transmission: for example, mutations in the haemagglutinin structure of influenza viruses limit their binding efficiency to the glycosylation of human sialic acid receptors, allowing prediction of the zoonotic potential of specific influenza A strains based on a few point mutations. Figure created with BioRender ([biorender.com](https://biorender.com))

availability will be dictated by each situation, but a balance should be struck between accessible biofluids (such as blood) and relevant tissues (such as lung tissue for respiratory viruses). On the other hand, given the ever-growing frequency of glycomic and glycoproteomic techniques (lectin-arrays or mass spectrometry-based), it is possible that the analytical work could be accomplished collaboratively at low to medium cost as part of broader comparative zoology and viral ecology. Critically, depositing the data sets generated by this work into existing public repositories like GlycoPOST and GlyYouCan, which is becoming standard practice in the glycobiology field (Rojas-Macias et al., 2019), will allow their value to grow exponentially for comparative and predictive research.

Recent viral epidemics and pandemics have highlighted a need for increased surveillance at the animal-human interface and forward planning of biochemical countermeasures (Rabozzi, 2020; Lurie et al., 2020; Munir et al., 2020). Characterizing the global glycome will help microbiologists and ecologists understand the broader dynamics of viral ecology, and these data could also easily be applied to understanding other pathogens. Moreover, as work during the COVID-19 pandemic has highlighted, understanding glycosylation as a viral phenotype is a key part of understanding pathogenesis and developing effective countermeasures (Shiliaev et al., 2021; Uraki & Kawaoka, 2021), and we suggest that building more glycomics into viral surveillance is a feasible, cost-effective, and impactful way to

expand the body of basic science that forms the basis for epidemic preparedness.

## AUTHOR CONTRIBUTIONS

Colin J. Carlson and Benjamin A. Neely conceptualized the perspective. All authors contributed to the writing and review of the present study.

## ACKNOWLEDGEMENTS

This work was supported by NSF BII 2021909 and NHMRC Ideas grant APP1186699. Identification of certain commercial equipment, instruments, software, or materials does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products identified are necessarily the best available for the purpose.

## CONFLICT OF INTEREST

The authors declare no competing interests.

## DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no data sets were generated or analysed during the current study.

## ORCID

Cassandra L. Pegg  <https://orcid.org/0000-0002-6080-7047>

Benjamin L. Schulz  <https://orcid.org/0000-0002-4823-7758>

Benjamin A. Neely  <https://orcid.org/0000-0001-6120-7695>

Colin J. Carlson  <https://orcid.org/0000-0001-6960-8434>

## REFERENCES

- Albery, G. F., Becker, D. J., Brierley, L., Brook, C. E., Christofferson, R. C., Cohen, L. E., Dallas, T. A., Eskew, E. A., Fagre, A., Farrell, M. J., Glennon, E., Guth, S., Joseph, M. B., Mollentze, N., Neely, B. A., Poisot, T., Rasmussen, A. L., Ryan, S. J., Seifert, S., ... Carlson, C. J. (2021). The science of the host-virus network. *Nature Microbiology*, 6(12), 1483–1492. <https://doi.org/10.1038/s41564-021-00999-5>
- Albery, G. F., Eskew, E. A., Ross, N., & Olival, K. J. (2020). Predicting the global mammalian viral sharing network using phylogeography. *Nature Communications*, 11(1), 2260. <https://doi.org/10.1038/s41467-020-16153-4>
- Alonso, J. M., Arroyuelo, A., Garay, P. G., Martin, O. A., & Vila, J. A. (2018). Finite dimension: A mathematical tool to analyse glycans. *Scientific Reports*, 8(1), 4426. <https://doi.org/10.1038/s41598-018-22575-4>
- Babayan, S. A., Orton, R. J., & Streicker, D. G. (2018). Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes. *Science*, 362(6414), 577–580. <https://doi.org/10.1126/science.aap9072>
- Bagdonaite, I., Vakhrushev, S. Y., Joshi, H. J., & Wandall, H. H. (2018). Viral glycoproteomes: Technologies for characterization and outlook for vaccine design. *FEBS Letters*, 592(23), 3898–3920. <https://doi.org/10.1002/1873-3468.13177>
- Bagdonaite, I., & Wandall, H. H. (2018). Global aspects of viral glycosylation. *Glycobiology*, 28(7), 443–467. <https://doi.org/10.1093/glycob/cwy021>
- Becker, D. J., Albery, G. F., Sjodin, A. R., Poisot, T., Bergner, L. M., Chen, B., Cohen, L. E., Dallas, T. A., Eskew, E. A., Fagre, A. C., Farrell, M. J., Guth, S., Han, B. A., Simmons, N. B., Stock, M., Teeling, E. C., & Carlson, C. J. (2022). Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. *Lancet Microbe*, 3(8), e625–e637. [https://doi.org/10.1016/S2666-5247\(21\)00245-7](https://doi.org/10.1016/S2666-5247(21)00245-7)
- Berndsen, Z. T., Chakraborty, S., Wang, X., Cottrell, C. A., Torres, J. L., Diedrich, J. K., Lopez, C. A., Yates, J. R., 3rd, van Gils, M. J., Paulson, J. C., Gnanakaran, S., & Ward, A. B. (2020). Visualization of the HIV-1 Env glycan shield across scales. *Proceedings of the National Academy of Sciences of the United States of America*, 117(45), 28014–28025. <https://doi.org/10.1073/pnas.2000260117>
- Burkholz, R., Quackenbush, J., & Bojar, D. (2021). Using graph convolutional neural networks to learn a representation for glycans. *Cell Reports*, 35(11), 109251. <https://doi.org/10.1016/j.celrep.2021.109251>
- Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., Fogarty, C. A., Barros, E. P., Taylor, B. C., McLellan, J. S., Fadda, E., & Amaro, R. E. (2020). Beyond shielding: The roles of glycans in the SARS-CoV-2 spike protein. *ACS Central Science*, 6(10), 1722–1734. <https://doi.org/10.1021/acscentsci.0c01056>
- Chothe, S. K., Bhushan, G., Nissly, R. H., Yeh, Y. T., Brown, J., Turner, G., Fisher, J., Sewall, B. J., Reeder, D. M., Terrones, M., Jayarao, B. M., & Kuchipudi, S. V. (2017). Avian and human influenza virus compatible sialic acid receptors in little brown bats. *Scientific Reports*, 7(1), 660. <https://doi.org/10.1038/s41598-017-00793-6>
- Clausen, T. M., Sandoval, D. R., Spliid, C. B., Pihl, J., Perrett, H. R., Painter, C. D., Narayanan, A., Majowicz, S. A., Kwong, E. M., McVicar, R. N., Thacker, B. E., Glass, C. A., Yang, Z., Torres, J. L., Golden, G. J., Bartels, P. L., Porell, R. N., Garretson, A. F., Laubach, L., ... Esko, J. D. (2020). SARS-CoV-2 infection depends on cellular heparan sulfate and ACE2. *Cell*, 183(4), 1043–1057 e1015. <https://doi.org/10.1016/j.cell.2020.09.033>
- Crocker, P. R., Paulson, J. C., & Varki, A. (2007). Siglecs and their roles in the immune system. *Nature Reviews. Immunology*, 7(4), 255–266. <https://doi.org/10.1038/nri2056>
- Cummings, R. D. (2009). The repertoire of glycan determinants in the human glycome. *Molecular BioSystems*, 5(10), 1087–1104. <https://doi.org/10.1039/b907931a>
- Di Giallonardo, F., Schlub, T. E., Shi, M., & Holmes, E. C. (2017). Dinucleotide composition in animal RNA viruses is shaped more by virus family than by host species. *Journal of Virology*, 91(8), e02381-02316. <https://doi.org/10.1128/JVI.02381-16>
- Dworkin, L. A., Clausen, H., & Joshi, H. J. (2022). Applying transcriptomics to study glycosylation at the cell type level. *iScience*, 25(6), 104419. <https://doi.org/10.1016/j.isci.2022.104419>
- Erikson, E., Wratil, P. R., Frank, M., Ambiel, I., Pahnke, K., Pino, M., Azadi, P., Izquierdo-Useros, N., Martinez-Picado, J., Meier, C., Schnaar, R. L., Crocker, P. R., Reutter, W., & Keppler, O. T. (2015). Mouse Siglec-1 mediates trans-infection of surface-bound Murine Leukemia virus in a sialic acid N-Acyl side chain-dependent manner. *The Journal of Biological Chemistry*, 290(45), 27345–27359. <https://doi.org/10.1074/jbc.M115.681338>
- Everest-Dass, A. V., Moh, E. S. X., Ashwood, C., Shathili, A. M. M., & Packer, N. H. (2018). Human disease glycomics: Technology advances enabling protein glycosylation analysis - part 1. *Expert Review of Proteomics*, 15(2), 165–182. <https://doi.org/10.1080/14789450.2018.1421946>
- Ferreira, R. C., Grant, O. C., Moyo, T., Dorfman, J. R., Woods, R. J., Travers, S. A., & Wood, N. T. (2018). Structural rearrangements maintain the glycan shield of an HIV-1 envelope trimer after the loss of a glycan. *Scientific Reports*, 8(1), 15031. <https://doi.org/10.1038/s41598-018-33390-2>
- Fischhoff, I. R., Castellanos, A. A., Rodrigues, J., Varsani, A., & Han, B. A. (2021). Predicting the zoonotic capacity of mammals to transmit SARS-CoV-2. *Proceedings of the Biological Sciences*, 288(1963), 20211651. <https://doi.org/10.1098/rspb.2021.1651>
- Gao, C., Wei, M., McKittrick, T. R., McQuillan, A. M., Heimburg-Molinaro, J., & Cummings, R. D. (2019). Glycan microarrays as chemical tools

- for identifying glycan recognition by immune proteins. *Frontiers in Chemistry*, 7, 833. <https://doi.org/10.3389/fchem.2019.00833>
- Ghorbani, M., Brooks, B. R., & Klauda, J. B. (2021). Exploring dynamics and network analysis of spike glycoprotein of SARS-CoV-2. *Biophysical Journal*, 120(14), 2902–2913. <https://doi.org/10.1016/j.bpj.2021.02.047>
- Han, B. A., Schmidt, J. P., Alexander, L. W., Bowden, S. E., Hayman, D. T., & Drake, J. M. (2016). Undiscovered bat hosts of filoviruses. *PLoS Neglected Tropical Diseases*, 10(7), e0004815. <https://doi.org/10.1371/journal.pntd.0004815>
- Hart, G. W., & Wells, L. (2021). Glycoproteomics: Making the study of the most structurally diverse and most abundant post-translational modifications more accessible to the scientific community. *Molecular & Cellular Proteomics: MCP*, 20, 100086. <https://doi.org/10.1016/j.mcpro.2021.100086>
- Jensen, P. H., Karlsson, N. G., Kolarich, D., & Packer, N. H. (2012). Structural analysis of N- and O-glycans released from glycoproteins. *Nature Protocols*, 7(7), 1299–1310. <https://doi.org/10.1038/nprot.2012.063>
- Jia, N., Byrd-Leotis, L., Matsumoto, Y., Gao, C., Wein, A. N., Lobby, J. L., Kohlmeier, J. E., Steinhauer, D. A., & Cummings, R. D. (2020). The human lung glycome reveals novel glycan ligands for influenza A virus. *Scientific Reports*, 10(1), 5320. <https://doi.org/10.1038/s41598-020-62074-z>
- Jones, J. E., Le Sage, V., & Lakdawala, S. S. (2021). Viral and host heterogeneity and their effects on the viral life cycle. *Nature Reviews. Microbiology*, 19(4), 272–282. <https://doi.org/10.1038/s41579-020-00449-9>
- Kasson, P. M., & Pande, V. S. (2008). Structural basis for influence of viral glycans on ligand binding by influenza hemagglutinin. *Biophysical Journal*, 95(7), L48–L50. <https://doi.org/10.1529/biophysj.108.141507>
- Kellman, B. P., & Lewis, N. E. (2021). Big-data glycomics: Tools to connect glycan biosynthesis to extracellular communication. *Trends in Biochemical Sciences*, 46(4), 284–300. <https://doi.org/10.1016/j.tibs.2020.10.004>
- Kolarich, D., Jensen, P. H., Altmann, F., & Packer, N. H. (2012). Determination of site-specific glycan heterogeneity on glycoproteins. *Nature Protocols*, 7(7), 1285–1298. <https://doi.org/10.1038/nprot.2012.062>
- Kuchipudi, S. V., Nelli, R., White, G. A., Bain, M., Chang, K. C., & Dunham, S. (2009). Differences in influenza virus receptors in chickens and ducks: Implications for interspecies transmission. *Journal of Molecular and Genetic Medicine: An International Journal of Biomedical Research*, 3(1), 143–151. <https://doi.org/10.4172/1747-0862.1000026>
- Lemmin, T., Soto, C., Stuckey, J., & Kwong, P. D. (2017). Microsecond dynamics and network analysis of the HIV-1 SOSIP Env trimer reveal collective behavior and conserved microdomains of the glycan shield. *Structure*, 25(10), 1631–1639. <https://doi.org/10.1016/j.str.2017.07.018>
- Li, Q., Xie, Y., Wong, M., Barboza, M., & Lebrilla, C. B. (2020). Comprehensive structural glycomic characterization of the glycolyxes of cells and tissues. *Nature Protocols*, 15(8), 2668–2704. <https://doi.org/10.1038/s41596-020-0350-4>
- Lurie, N., Saville, M., Hatchett, R., & Halton, J. (2020). Developing Covid-19 vaccines at pandemic speed. *The New England Journal of Medicine*, 382(21), 1969–1973. <https://doi.org/10.1056/NEJMp2005630>
- Maginnis, M. S. (2018). Virus-Receptor Interactions: The key to cellular invasion. *Journal of Molecular Biology*, 430(17), 2590–2611. <https://doi.org/10.1016/j.jmb.2018.06.024>
- McVeigh, P., Cwiklinski, K., Garcia-Campos, A., Mulcahy, G., O'Neill, S. M., Maule, A. G., & Dalton, J. P. (2018). In silico analyses of protein glycosylating genes in the helminth *Fasciola hepatica* (liver fluke) predict protein-linked glycan simplicity and reveal temporally-dynamic expression profiles. *Scientific Reports*, 8(1), 11700. <https://doi.org/10.1038/s41598-018-29673-3>
- Mollentez, N., Babayan, S. A., & Streicker, D. G. (2021). Identifying and prioritizing potential human-infecting viruses from their genome sequences. *PLoS Biology*, 19(9), e3001390. <https://doi.org/10.1371/journal.pbio.3001390>
- Moremen, K. W., Tiemeyer, M., & Nairn, A. V. (2012). Vertebrate protein glycosylation: Diversity, synthesis and function. *Nature Reviews. Molecular Cell Biology*, 13(7), 448–462. <https://doi.org/10.1038/nrm3383>
- Munir, K., Ashraf, S., Munir, I., Khalid, H., Muneer, M. A., Mukhtar, N., Amin, S., Ashraf, S., Imran, M. A., Chaudhry, U., Zaheer, M. U., Arshad, M., Munir, R., Ahmad, A., & Zhao, X. (2020). Zoonotic and reverse zoonotic events of SARS-CoV-2 and their impact on global health. *Emerging Microbes & Infections*, 9(1), 2222–2235. <https://doi.org/10.1080/22221751.2020.1827984>
- Newhouse, E. I., Xu, D., Markwick, P. R., Amaro, R. E., Pao, H. C., Wu, K. J., Alam, M., McCammon, J. A., & Li, W. W. (2009). Mechanism of glycan receptor recognition and specificity switch for avian, swine, and human adapted influenza virus hemagglutinins: A molecular dynamics perspective. *Journal of the American Chemical Society*, 131(47), 17430–17442. <https://doi.org/10.1021/ja904052q>
- Nguyen, L., McCord, K. A., Bui, D. T., Bouwman, K. M., Kitova, E. N., Elaish, M., Kumawat, D., Daskhan, G. C., Tomris, I., Han, L., Chopra, P., Yang, T. J., Willows, S. D., Mason, A. L., Mahal, L. K., Lowary, T. L., West, L. J., Hsu, S. D., Hobman, T., ... Klassen, J. S. (2021). Sialic acid-containing glycolipids mediate binding and viral entry of SARS-CoV-2. *Nature Chemical Biology*, 18(1), 81–90. <https://doi.org/10.1038/s41589-021-00924-1>
- Ohtsubo, K., & Marth, J. D. (2006). Glycosylation in cellular mechanisms of health and disease. *Cell*, 126(5), 855–867. <https://doi.org/10.1016/j.cell.2006.08.019>
- Oliveira, T., Thaysen-Andersen, M., Packer, N. H., & Kolarich, D. (2021). The Hitchhiker's guide to glycoproteomics. *Biochemical Society Transactions*, 49(4), 1643–1662. <https://doi.org/10.1042/BST20200879>
- Park, Y. J., Walls, A. C., Wang, Z., Sauer, M. M., Li, W., Tortorici, M. A., Bosch, B. J., DiMaio, F., & Veelsler, D. (2019). Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors. *Nature Structural & Molecular Biology*, 26(12), 1151–1157. <https://doi.org/10.1038/s41594-019-0334-7>
- Rabozzi, G. (2020). Emerging zoonoses: A one health challenge. *EclinicalMedicine*, 19, 100300. <https://doi.org/10.1016/j.eclinm.2020.100300>
- Raman, R., Tharakaraman, K., Sasisekharan, V., & Sasisekharan, R. (2016). Glycan-protein interactions in viral pathogenesis. *Current Opinion in Structural Biology*, 40, 153–162. <https://doi.org/10.1016/j.sbi.2016.10.003>
- Remoroza, C. A., Burke, M. C., Liu, Y., Mirokhin, Y. A., Tchekhovskoi, D. V., Yang, X., & Stein, S. E. (2021). Representing and comparing site-specific glycan abundance distributions of glycoproteins. *Journal of Proteome Research*, 20(9), 4475–4486. <https://doi.org/10.1021/acs.jproteome.1c00442>
- Riley, N. M., Bertozzi, C. R., & Pitteri, S. J. (2020). A pragmatic guide to enrichment strategies for mass spectrometry-based glycoproteomics. *Molecular & Cellular Proteomics: MCP*, 20, 100029. <https://doi.org/10.1074/mcp.R120.002277>
- Rojas-Macias, M. A., Mariethoz, J., Andersson, P., Jin, C., Venkatakrisnan, V., Aoki, N. P., Shinmachi, D., Ashwood, C., Madunic, K., Zhang, T., Miller, R. L., Horlacher, O., Struwe, W. B., Watanabe, Y., Okuda, S., Levander, F., Kolarich, D., Rudd, P. M., Wuhrer, M., ... Karlsson, N. G. (2019). Towards a standardized bioinformatics infrastructure for N- and O-glycomics. *Nature Communications*, 10(1), 3275. <https://doi.org/10.1038/s41467-019-11131-x>

- Schjoldager, K. T., Narimatsu, Y., Joshi, H. J., & Clausen, H. (2020). Global view of human protein glycosylation pathways and functions. *Nature Reviews. Molecular Cell Biology*, 21(12), 729–749. <https://doi.org/10.1038/s41580-020-00294-x>
- Seitz, C., Casalino, L., Konecny, R., Huber, G., Amaro, R. E., & McCammon, J. A. (2020). Multiscale simulations examining glycan shield effects on drug binding to influenza neuraminidase. *Biophysical Journal*, 119(11), 2275–2289. <https://doi.org/10.1016/j.bpj.2020.10.024>
- Shackelton, L. A., Parrish, C. R., & Holmes, E. C. (2006). Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *Journal of Molecular Evolution*, 62(5), 551–563. <https://doi.org/10.1007/s00239-005-0221-1>
- Shiliaev, N., Lukash, T., Palchevska, O., Crossman, D. K., Green, T. J., Crowley, M. R., Frolova, E. I., & Frolov, I. (2021). Natural and recombinant SARS-CoV-2 isolates rapidly evolve in vitro to higher infectivity through more efficient binding to heparan sulfate and reduced S1/S2 cleavage. *Journal of Virology*, 95(21), e0135721. <https://doi.org/10.1128/JVI.01357-21>
- Shinya, K., Ebina, M., Yamada, S., Ono, M., Kasai, N., & Kawaoka, Y. (2006). Avian flu: influenza virus receptors in the human airway. *Nature*, 440(7083), 435–436. <https://doi.org/10.1038/440435a>
- Short, K. R., Richard, M., Verhagen, J. H., van Riel, D., Schrauwen, E. J., van den Brand, J. M., Manz, B., Bodewes, R., & Herfst, S. (2015). One health, multiple challenges: The inter-species transmission of influenza A virus. *One Health (Amsterdam, Netherlands)*, 1, 1–13. <https://doi.org/10.1016/j.onehlt.2015.03.001>
- Spiro, R. G. (2002). Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, 12(4), 43R–56R. <https://doi.org/10.1093/glycob/12.4.43r>
- Stevens, J., Blixt, O., Glaser, L., Taubenberger, J. K., Palese, P., Paulson, J. C., & Wilson, I. A. (2006). Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *Journal of Molecular Biology*, 355(5), 1143–1155. <https://doi.org/10.1016/j.jmb.2005.11.002>
- Stewart-Jones, G. B., Soto, C., Lemmin, T., Chuang, G. Y., Druz, A., Kong, R., Thomas, P. V., Wagh, K., Zhou, T., Behrens, A. J., Bylund, T., Choi, C. W., Davison, J. R., Georgiev, I. S., Joyce, M. G., Kwon, Y. D., Pancera, M., Taft, J., Yang, Y., ... Kwong, P. D. (2016). Trimeric HIV-1-Env structures define glycan shields from clades A, B, and G. *Cell*, 165(4), 813–826. <https://doi.org/10.1016/j.cell.2016.04.010>
- Suenaga, T., & Arase, H. (2015). Viral interactions with glycans. In N. Taniguchi, T. Endo, G. Hart, P. Seeberger, & C. H. Wong (Eds.), *Glycoscience: Biology and Medicine*. Tokyo: Springer. [https://doi.org/10.1007/978-4-431-54841-6\\_152](https://doi.org/10.1007/978-4-431-54841-6_152)
- Sztain, T., Ahn, S. H., Bogetti, A. T., Casalino, L., Goldsmith, J. A., Seitz, E., McCool, R. S., Kearns, F. L., Acosta-Reyes, F., Maji, S., Mashayekhi, G., McCammon, J. A., Ourmazd, A., Frank, J., McLellan, J. S., Chong, L. T., & Amaro, R. E. (2021). A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nature Chemistry*, 13(10), 963–968. <https://doi.org/10.1038/s41557-021-00758-3>
- Takata, M. A., Goncalves-Carneiro, D., Zang, T. M., Soll, S. J., York, A., Blanco-Melo, D., & Bieniasz, P. D. (2017). CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature*, 550(7674), 124–127. <https://doi.org/10.1038/nature24039>
- Thompson, A. J., de Vries, R. P., & Paulson, J. C. (2019). Virus recognition of glycan receptors. *Current Opinion in Virology*, 34, 117–129. <https://doi.org/10.1016/j.coviro.2019.01.004>
- Thompson, A. J., & Paulson, J. C. (2021). Adaptation of influenza viruses to human airway receptors. *The Journal of Biological Chemistry*, 296, 100017. <https://doi.org/10.1074/jbc.REV120.013309>
- Uraki, R., & Kawaoka, Y. (2021). Host glycolipids in SARS-CoV-2 entry. *Nature Chemical Biology*, 18(1), 6–7. <https://doi.org/10.1038/s41589-021-00923-2>
- Varki, A. (2017). Biological roles of glycans. *Glycobiology*, 27(1), 3–49. <https://doi.org/10.1093/glycob/cww086>
- Watanabe, Y., Bowden, T. A., Wilson, I. A., & Crispin, M. (2019). Exploitation of glycosylation in enveloped virus pathobiology. *Biochimica et Biophysica Acta - General Subjects*, 1863(10), 1480–1497. <https://doi.org/10.1016/j.bbagen.2019.05.012>
- Werz, D. B., Ranzinger, R., Herget, S., Adibekian, A., von der Lieth, C. W., & Seeberger, P. H. (2007). Exploring the structural diversity of mammalian carbohydrates (“glycospace”) by statistical databank analysis. *ACS Chemical Biology*, 2(10), 685–691. <https://doi.org/10.1021/cb700178s>
- West, C. M., Malzl, D., Hykollari, A., & Wilson, I. B. H. (2021). Glycomics, glycoproteomics, and glycogenomics: An inter-taxa evolutionary perspective. *Molecular & Cellular Proteomics: MCP*, 20, 100024. <https://doi.org/10.1074/mcp.R120.002263>
- Wood, N. T., Fadda, E., Davis, R., Grant, O. C., Martin, J. C., Woods, R. J., & Travers, S. A. (2013). The influence of N-linked glycans on the molecular dynamics of the HIV-1 gp120 V3 loop. *PLoS One*, 8(11), e80301. <https://doi.org/10.1371/journal.pone.0080301>
- Xu, D., Newhouse, E. I., Amaro, R. E., Pao, H. C., Cheng, L. S., Markwick, P. R., McCammon, J. A., Li, W. W., & Arzberger, P. W. (2009). Distinct glycan topology for avian and human sialopentasaccharide receptor analogues upon binding different hemagglutinins: a molecular dynamics perspective. *Journal of Molecular Biology*, 387(2), 465–491. <https://doi.org/10.1016/j.jmb.2009.01.040>
- Yang, M., Huang, J., Simon, R., Wang, L. X., & MacKerell, A. D., Jr. (2017). Conformational heterogeneity of the HIV envelope glycan shield. *Scientific Reports*, 7(1), 4435. <https://doi.org/10.1038/s41598-017-04532-9>
- Yang, Q., Hughes, T. A., Kelkar, A., Yu, X., Cheng, K., Park, S., Huang, W. C., Lovell, J. F., & Neelamegham, S. (2020). Inhibition of SARS-CoV-2 viral entry upon blocking N- and O-glycan elaboration. *eLife*, 9, e61552. <https://doi.org/10.7554/eLife.61552>
- Zhao, J., Song, E., Huang, Y., Yu, A., & Mechref, Y. (2021). Variability in the glycosylation patterns of gp120 proteins from different human immunodeficiency virus type 1 isolates expressed in different host cells. *Journal of Proteome Research*, 20(10), 4862–4874. <https://doi.org/10.1021/acs.jproteome.1c00587>
- Zhao, P., Praissman, J. L., Grant, O. C., Cai, Y., Xiao, T., Rosenbalm, K. E., Aoki, K., Kellman, B. P., Bridger, R., Barouch, D. H., Brindley, M. A., Lewis, N. E., Tiemeyer, M., Chen, B., Woods, R. J., & Wells, L. (2020). Virus-receptor interactions of glycosylated SARS-CoV-2 spike and human ACE2 receptor. *Cell Host & Microbe*, 28(4), 586–601 e586. <https://doi.org/10.1016/j.chom.2020.08.004>

**How to cite this article:** Pegg, C. L., Schulz, B. L., Neely, B. A., Albery, G. F., & Carlson, C. J. (2022). Glycosylation and the global virome. *Molecular Ecology*, 00, 1–8. <https://doi.org/10.1111/mec.16731>