University of Southampton

Faculty of Environmental and Life Sciences

School of Biological Sciences

The glycan shields of HIV-1 and SARS-CoV-2 spike proteins and their differential importance in vaccine design

by

Joel D. Allen

ORCID ID: 0000-0003-2547-968X

Thesis for the degree of Doctor of Philosophy

July 2021

University of Southampton

Abstract

Viral infections are responsible for major global pandemics, causing millions of deaths and greatly impacting human societies across the globe. Two of the viruses responsible for some of the worst pandemics over the last century are human immunodeficiency virus-1 (HIV-1) and severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). Whilst symptoms of the diseases they cause are distinct, there are features of the virions which are shared across both viruses. Proteins on the surface of the virion mediate interactions with host receptors and act as fusion machinery to enable the virus to enter the host cell. As these are exposed, they are key targets of the human immune system, particularly for neutralizing antibodies. The elicitation of broadly neutralizing antibodies is an attractive target for vaccination as these antibodies, in combination with other parts of the immune system, protect against viral exposure.

Whilst the immune system is able to target the receptor binding proteins on the virion, viruses employ a range of strategies to avoid clearance. One such strategy is to shield the antigenic viral protein surface with glycans. During egress through the host cell, spike proteins hijack the host's post translational machinery and attach polysaccharide chains, known as glycans, which cover the protein surface. As these are host-derived it can be difficult for the immune system to recognize these glycans as foreign and can aid the virus in avoiding immune detection. Vaccine design must capture the native-like display of glycans when trying to elicit an antibody-based response so as to avoid the recognition of vaccine-specific epitopes that are not present on the virus. The attachment of glycans to viral spikes is a heterogeneous process, as glycan maturation is not genetically encoded, as it is for amino acids. To this end bespoke strategies are required to understand glycan presentation on viruses and their corresponding vaccine candidates. Mass spectrometry is ideal for studying glycans as it can handle extremely heterogeneous mixtures. In this thesis I demonstrate how mass spectrometry can be used to inform vaccine design efforts by comparing and contrasting the glycosylation of SARS-CoV-2 and HIV-1 spike proteins, by studying the glycosylation of leading vaccine candidates for both viruses and their viral counterparts.

This thesis demonstrates that whilst key regions of the HIV-1 Env glycan shield are conserved between recombinant protein and the corresponding virus, there are differences in glycosylation regarding the composition and glycan occupancy at particular regions of the glycan shield. These differences result in the elicitation of undesirable antibodies however, this thesis demonstrates that the glycan occupancy of recombinant Env can be improved by engineering the N-linked glycan attachment sequon. Comparable analyses of SARS-CoV-2 S protein revealed that whilst under processed oligomannose-type glycans are present on the spike glycoprotein, their abundance is much less than that of HIV-1 Env, indicating a less dense glycan shield. Additionally, the glycan shield of recombinant and viral-derived S protein were in closer agreement than observed for HIV-1 Env. This work indicates that the extensive glycan shield of HIV-1 is more of a hinderance to vaccine design efforts compared to SARS-CoV-2.

Table of Contents

Table of Tables

Table of Figures

Table of Figures

Research Thesis: Declaration of Authorship

Print name: **Joel D Allen**

Title of thesis: **The glycan shields in HIV-1 and SARS-CoV-2 and their differential importance in vaccine design**

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University.
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
3. Where I have consulted the published work of others, this is always clearly attributed.
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
5. I have acknowledged all main sources of help.
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.
7. Parts of this work have been published as:-

   Chapter 1: Allen, J. D., Sanders, R. W., Doores, K. J. & Crispin, M. Harnessing post-translational modifications for next-generation HIV immunogens. Biochem. Soc. Trans. 46, 691–698 (2018).
   Chapter 3: Struwe, W. B., Chertova, E, Allen, J. D., et al. Site-Specific Glycosylation of Virion-Derived HIV-1 Env Is Mimicked by a Soluble Trimeric Immunogen. Cell Rep. 24, 1958-1966.e5 (2018).
   Chapter 4: Derking, R., Allen, J. D., *et al*. Enhancing glycan occupancy of soluble HIV-1 envelope trimers to mimic the native viral spike. Cell Rep. 35, 108933 (2021).
   Chapter 5: Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. Science 369, 330–333 (2020).
   Chapter 6: 1. Allen, J. D. et al. Site-Specific Steric Control of SARS-CoV-2 Spike Glycosylation. Biochemistry acs.biochem.1c00279 (2021) doi:10.1021/acs.biochem.1c00279.
   Chapter 7: Allen, J. D., Watanabe, Y., Chawla, H., Newby, M. L. & Crispin, M. Subtle Influence of ACE2 Glycan Processing on SARS-CoV-2 Recognition. J. Mol. Biol. 433, 166762 (2021).

Signature:  Date: 19/07/2021

Acknowledgements

There are so many people to thank, without who this thesis would not be possible. Firstly, thanks to Snezana for guiding me through the early days of my research and putting up with all the many mistakes I made. In addition, a big thanks goes to Gemma and Yasunori for helping out with every part of my PhD, its mad how much stuff has happened over the last few years, including packing up the entire lab and moving to Southampton.

Gemma deserves a special mention for tolerating me for an entire year without anyone else working in the wet lab, no easy feat. Also a major part of the work in this thesis was undertaken during the COVID-19 lockdowns and special thanks go to Yasunori and Himanshi, it feels like for a good few months you were the only people who I saw face to face, and the fact that none of us ended up hating each other is a miracle.

Also, thanks to the whole Southampton crew, especially Maddy who has kept me sane throughout the whole process, you're the best! Also, Al, Abi, Dilys, Lydia, Hannah, John, Charlie E, Charlie B, Anna, Lan, Weston, Re'em and Li Chieh couldn't ask for better lab mates, hopefully it's been as enjoyable for you guys as it has for me.

Thanks to Max Crispin for supervising me all these years, the stuff I've learnt so far will definitely help me out for years to come, I appreciate the opportunity and I'm glad you managed to find the funds to keep me on as a PhD student.

Additionally, it's been great working with labs from all over the world, big thanks to Ronald, definitely come back and visit the UK again soon, maybe when the Netherlands have a half decent football team. Thanks to everyone at Scripps who showed me around, you guys have been great to collaborate with over the years, especially Raiees and Kimmo. Also, thanks to all the collaborators who made this thesis possible; notably the laboratories of Prof. Lifson, Prof. Sanders, Prof. McLellan, and Prof. Peter Bond.

Finally, thanks to the Allen's, especially Mum and Dad, I said I'd get to the end eventually!

Definitions and Abbreviations

2-AA.....2 aminobenzoic acid; a label for released glycans to enable quantification.

2G12.....A broadly neutralizing anti-HIV-1 antibody that targets glycans on the envelope glycoprotein.

SARS-CoV-2 2P S.....Stabilising mutations introduced into the SARS-CoV-2 spike protein which stabilises S protein in the pre-fusion conformation.

ACE2 .....Angiotensin converting enzyme 2.

AIDS .....Acquired immunodeficiency syndrome.

APC .....Antigen presenting cells

ASA.....Accessible surface area

BG505 SOSIP.664 .....Stabilising mutations introduced into the HIV-1 Envelope glycoprotein from the BG505 clade A/E isolate which maintains a native-like conformation as a soluble protein.

BG505 SOSIP.664.v4.1 ..Additional stabilising mutations introduced into BG505 SOSIP.664 to further increase SOSIP stability.

BG505 SOSIP.664-GT1..A glycan depleted version of BG505 SOSIP.664 to activate B cells producing antibodies that are the precursors to broadly neutralising antibodies

BG505.T332N-LAI .....BG505 Env attached to a clade B lab adapted isolated (LAI) backbone, containing BG505 Env with C-terminus and transmembrane domains derived from LAI.

bnAb .....Broadly neutralizing antibody

BN-PAGE.....Blue Native polyacrylamide gel electrophoresis

C18.....Octadecyl carbon chain

CCR5.....C-C chemokine receptor type 5

CD4 .....Cluster of differentiation 4

CD4bs.....CD4 binding site

CD8 .....Cluster of differentiation 8

ChAdOx-nCoV-19.....The Oxford–AstraZeneca COVID-19 vaccine, a viral vector vaccine for prevention of COVID-19.

CHO ..... Chinese Hamster Ovary cells.

CID ..... Collision induced dissociation

COVID-19 ..... Coronavirus Disease- 2019

Cryo-EM ..... Cryo-electron microscopy

CXCR4 ..... C-X-C chemokine receptor type 4

DTT ..... Dithiothreitol

E protein ..... Envelope protein

Easy-nLC 1200 ..... Nanoflow liquid chromatography instrument used for the separation of glycopeptides prior to analysis by mass spectrometry.

EndoH ..... Endoglycosidase H

Env ..... Envelope glycoprotein of HIV-1

eOD ..... Engineered outer domain

ER ..... Endoplasmic reticulum

ERGIC ..... ER/Golgi intermediate compartment

ESI ..... Electrospray ionization

Fc ..... Fragment crystallizable, the tail region of immunoglobulins mediating effector functions with the immune system.

Gal ..... Galactose

GALT ..... Gut-associated lymphoid tissue

Glc ..... Glucose

GlcNAc ..... N-acetylglucosamine

GMP ..... Good Manufacturing Practise

GnTI ..... N-acetylglucosaminyltransferase I

gp120 ..... Receptor binding portion of HIV-1 Env

gp140 ..... HIV-1 Env glycoprotein with a truncated C-terminus to solubilize the protein.

gp160 ..... Precursor polypeptide of HIV-1 Env that exists prior to furin cleavage.

gp41 ..... Portion of HIV-1 Env which is responsible for membrane fusion.

HA..... Influenza Haemagglutinin

HCD..... Higher energy collision-induced dissociation

HEK293F ..... Human embryonic kidney 293 Freestyle cells, used for recombinant protein production.

Hex ..... Hexose

HexNAc..... N-Acetylhexosamine

HILIC-UPLC..... Ultra-high-performance liquid chromatography

HIV-1 ..... Human immunodeficiency virus-1

HPLC ..... High performance liquid chromatography

HRV3C..... Human rhinovirus 3C protease

HxB2..... HIV-1 reference strain used for alignment of the diverse HIV-1 sequences.

IAA..... Iodoacetamide

IgG..... Immunoglobulin G

iGL..... Inferred germline

IMC..... Infectious molecular clone

IM-MS..... Ion-mobility tandem mass spectrometry

IMP ..... Intrinsic mannose patch

IN..... HIV-1 integrase

K_D ..... Dissociation constant

Kif ..... Kifunensine

LASV GPC ..... Lassa virus glycoprotein complex

LC-MS ..... Liquid chromatography-mass spectrometry

M protein..... SARS-CoV-2 membrane protein

Man..... Mannose

MERS-CoV ..... Middle Eastern respiratory syndrome coronavirus

MHC ..... Major histocompatibility complex

mRNA ..... Messenger ribonucleic acid

MS-2 or MSMS ..... Mass spectrometry terminology referring to the process of isolating and fragmenting precursor ions to enable structural characterisation of an ion of interest.

N protein..... Nucleoprotein of SARS-CoV-2

NeuAc..... Neuraminic acid

NFL ..... Native flexibly linked

NS-EM..... Negative stain- electron microscopy

NTD..... N-terminal domain

NxS/T ..... Amino acid sequence that defines site of attachment for an N-linked glycan.

ORFs..... Open reading frames

PDB..... Protein database

PDI ..... Protein disulphide isomerase

PEImax ..... Polyethylenimine "Max" reagent used for transient transfection of cells.

PNGaseF..... Peptide N-glycosidase F

PNGS ..... Potential N-linked glycosylation site

PRR..... Pattern recognition receptor

PVDF..... Polyvinylidene fluoride

RBD..... SARS-CoV-2 receptor binding domain.

RdRp ..... RNA replicase–transcriptase complex

RM20E1 ..... Antibody isolated from a macaque, which can only neutralize virus lacking the N611 glycan site.

RT ..... HIV-1 reverse transcriptase

S protein..... SARS-CoV-2 Spike glycoprotein

SARS-CoV-1..... Severe acute respiratory syndrome coronavirus-1

SARS-CoV-2..... severe acute respiratory syndrome coronavirus-2

SDS-PAGE ..... Sodium dodecyl sulphate–polyacrylamide gel electrophoresis

SEC..... Size exclusion chromatography

Definitions and Abb

SPR ..... Surface plasmon resonance

ST6Gal1 (ST6) ..... Beta-galactoside alpha-2,6-sialyltransferase 1

STT3A /B ..... Catalytic subunit of the oligosaccharyl transferase (OST) complex that catalyses the initial transfer of a glycan to a polypeptide.

TAMP..... Trimer associated mannose patch

TCR ..... T cell receptor

TMPRSS2..... Transmembrane protease, serine 2

V1/V2..... Variable loops 1 and 2 located at the apex of HIV-1 Env

WHO..... World health organization

WT ..... Wild type

XIC..... Extracted ion chromatogram

Chapter 1 Introduction

1.1 Pandemic viruses and their impact on the human population

Over the last century several major global pandemics have caused millions of deaths worldwide. The Spanish Flu outbreak in the early 20th Century is thought to have killed an estimated 50 million people, for example [1]. In the 21st century the emergence of a novel coronavirus from Wuhan, China, rapidly spread across the globe [2]. This virus was termed SARS-CoV-2, the causative agent of COVID-19. Its infectivity and mortality rates in elderly populations has meant that governments worldwide have had to impose extensive measures to try to control this virus, which has meant that the COVID-19 pandemic has had a much wider reaching impact beyond those who have contracted the virus. Throughout 2021, the impact of rapid vaccine deployment demonstrates the critical role that immunogen design efforts continue to play and that continued advances in vaccine development are needed to ensure vaccine candidates can be rapidly deployed in future pandemics [3]. There have been a number of successful vaccines developed which provide high levels of protection from the original Wuhan SARS-CoV-2 virus and corresponding variants, and the development of these vaccines was aided by advances in vaccine design catalysed by the search for a vaccine against HIV-1 [4]. The AIDS pandemic has impacted millions of lives and remains one of the leading causes of death globally and, importantly, in some regions of Sub-Saharan Africa is the leading cause of death [5]. Whilst many options for suppression and prophylaxis are now available, a vaccine that can protect against infection remains a key goal for the global scientific community. Vaccine design efforts against HIV-1 have battled against an extremely difficult target, due to the extensive mutation rates and the ability for the virus to lay dormant within cells [6,7]. This has meant that novel innovations have been required to begin to move towards a vaccine, including new vaccination strategies that previously did not exist.

Despite differences in viral life cycles, pathogenesis and symptoms, the target of vaccine design efforts against both diseases is similar. For both viruses, the receptor binding envelope glycoproteins are the primary target for immunogen design efforts due to their exposure to the immune system on the surface of infectious virions. Antibodies elicited against viral envelope glycoproteins can neutralize infectivity by blocking receptor binding, and therefore defining the antigenic surface of viral spike glycoproteins is important as it aids with the identification of potential neutralizing antibody epitopes. Another conserved feature of viruses such as HIV-1 and SARS-CoV-2 is the presence of amino acid sequences

that act as acceptors for the post-translational attachment of glycans. Viral glycoproteins possess high numbers of these sequons, and the attachment of large N-linked glycans across the envelope glycoproteins forms a shield, whereby the underlying protein is occluded from the humoral immune system. This means that understanding the composition and abundance of glycans covering the surface of the envelope glycoprotein is an intrinsic part of defining the antigenic surface available for immune recognition.

As obligate parasites, viruses require the host cell to perform post-translational modifications on viral proteins and therefore the processes employed by both SARS-CoV-2 and HIV-1 will converge. This thesis aims to explore how the glycan shields of two major pandemic viruses; HIV-1 and SARS-CoV-2 compare, using a range of biophysical techniques, focused on mass spectrometry. For both viruses, the integrity of the glycan shield will be investigated by comparing viral-derived envelope glycoproteins with recombinant, solubilized variants, which form the basis of immunogen design efforts. These distinct examples will be used to establish potential global phenomena in how viral glycan shields are formed, and then to discuss the differential importance of glycan shields in vaccine design, and why these differences have arisen. To explore the factors that influence the presentation of glycans on envelope glycoproteins a knowledge of the viral life cycle and its pathogenesis is required. In addition, the mechanisms that the host cell uses to perform the post-translational attachment and processing of glycans is necessary. Throughout **Chapter 1**, this thesis aims to explore the factors that influence the presentation of glycans on envelope glycoproteins and then demonstrate how innovations in the study of the HIV-1 glycan shield have been rapidly deployed to investigate SARS-CoV-2.

1.2 Acquired immunodeficiency syndrome and the global impact of HIV infection

1.2.1 The emergence of human immunodeficiency virus-1 (HIV-1)

Acquired immune deficiency syndrome (AIDS) was first described in 1981 when five cases were reported of *Pneumocystis. pneumonia* in previously healthy men in Los Angeles [8] and eight cases of Kaposi's Sarcoma, a cancer normally affecting older populations, was observed in younger men in New York [9]. Since its characterisation, infections from the human-immunodeficiency virus (HIV) have been linked to 34 million deaths worldwide, with approximately 36.9 million people living with HIV and 1.8 million people becoming newly infected with HIV in 2016 [10]. HIV is a member of the Lentivirus genus in the Retroviridae family. The lentivirus genus is characterized by a long incubation period of infection and the ability to infect non-dividing cells [11]. As a retrovirus, the viral genome of HIV consists of RNA. Upon infection of a host cell this RNA is reverse transcribed to DNA and inserted into the host cell genome using an integrase enzyme. The infected host cell then translates viral proteins resulting in the formation of more viral particles. The immune deficiency from HIV infection arises because the virus infects and kills CD4+ T cells, reducing the CD4+ T cell count over a number of years [12]. If left untreated, these levels fall below a certain threshold and cell-mediated immunity is compromised. Opportunistic infections, such as of *Pneumocystis pneumonia*, are fatal in immunocompromised patients whereas healthy individuals' immune systems clear such infections without severe symptoms developing.

Even though AIDS is such a prominent disease worldwide, a vaccine that provides broad protection against HIV infection is yet to be developed. One example of a vaccine that has reached phase 3 clinical trials only had a 30% efficacy [13]. A trial developed on a similar platform was terminated early due to a lack of efficacy [14]. This can be attributed, in part, to the presence of many quasispecies of HIV that exist in an infected host [15]. It is estimated that the number of different quasispecies in a single infected patient is comparable to the total genetic diversity of the Influenza virus worldwide [16]. This incredible genetic diversity is a result of the retrovirus utilising an error-prone reverse transcriptase to convert the viral RNA into DNA. As a result, protein epitopes are not conserved and any approach that exploits solely protein epitopes will result in a selective pressure that is quickly overcome by the rapidly mutating HIV population.

The animal reservoir for HIV-1 is likely immunodeficiency viruses found in wild chimpanzees (SIVcpz) [17], and likely crossed over multiple times from chimpanzees to the human population [18]. The majority of circulating HIV-1 viruses are classified into the M-

group (main) which is further subdivided into clades A, B, C, D, F, G, H, and J in addition to circulating recombinant forms (CRFs) [19]. Different clades predominate across different regions of the globe, with clade B prevalent in Europe and America, clade C in South Africa and clades C, A and D in sub-Saharan Africa. As such a vaccine will need to induce protection against all clades circulating worldwide [20].

1.2.2 **The viral genome and life cycle of HIV-1**

The HIV-1 genome encodes for a number of structural and non-structural proteins which perform a variety of roles to aid in viral replication. The virus contains two copies of single stranded positive sense RNA, which encodes for 9 genes (*gag*, *pol*, *env*, *tat*, *rev*, *nef*, *vif*, *vpr* and *vpu*) (**Figure 1-1**) [21]. Viral proteins are encoded for by three of these genes; *gag*, *pol* and *env*. The *gag* gene directs the synthesis, transport to the plasma membrane, and assembly of the structural precursor polyprotein Gag, resulting in formation of viral particles [22]. The *pol* gene encodes for key enzymes in the viral cycle, such as protease (PR), reverse transcriptase (RT) and integrase (IN) and the *env* gene encodes for the Envelope glycoprotein (Env), which is translated as a single gp160 polypeptide. This polypeptide is cleaved by the host protease, furin, during ER/Golgi transit into gp120 and gp41, which form a dimer. These heterodimers assemble as trimers on the mature virion.

When an infectious HIV-1 particle encounters a CD4+ cell the gp120 portion of Env binds. In addition to CD4 a co-receptor is required for binding. In the case of HIV-1 this can either be C-C chemokine receptor type 5 (CCR5) or C-X-C chemokine receptor type 4 (CXCR4) [23]. Different viruses can have tropism for either co-receptor and are referred to as R4 or R5 tropic viruses. The key role of these co-receptors for HIV-1 infection were evident to the broader community when the first person was “cured” of HIV-1 infection. This patient was receiving treatment for cancer and, following suppression of their own immune system, received a stem cell transplant with cells that did not express CCR5 and remission was reported 18 months following treatment [24]. Following receptor binding Env undergoes conformational change and the gp120s shed, allowing the gp41 portion to extend and expose its fusion peptide, which then inserts into the plasma membrane [25]. This results in membrane fusion and the internal components of the virus enter the cell. Once inside the cell, the single stranded +ve sense RNA is reverse transcribed to DNA by the RT enzyme. This is the major source of the genetic variability of HIV-1, as this enzyme is extremely error prone [26]. Following this process, the DNA migrates to the nucleus, along with integrase. The integrase then incorporates the viral DNA into the host genome [27]. This process is what enables HIV-1 to become a latent infection. Anti-retroviral treatments can prevent the production of more infectious particles, but most are not able to remove the integrated DNA from the genome, and so HIV-1 latency prevents clearance of infection using anti-virals [28].

The virus then uses the transcription and translation machinery of the host to produce new infectious virions with proteins encoded by *gag* and *pol* being translated into the cytoplasm, whereas the Env protein enters the secretory machinery of the ER and Golgi, undergoing post-translational modification as would a normal cell-surface glycoprotein [29]. The virions assemble at the plasma membrane and then spread to other cells.

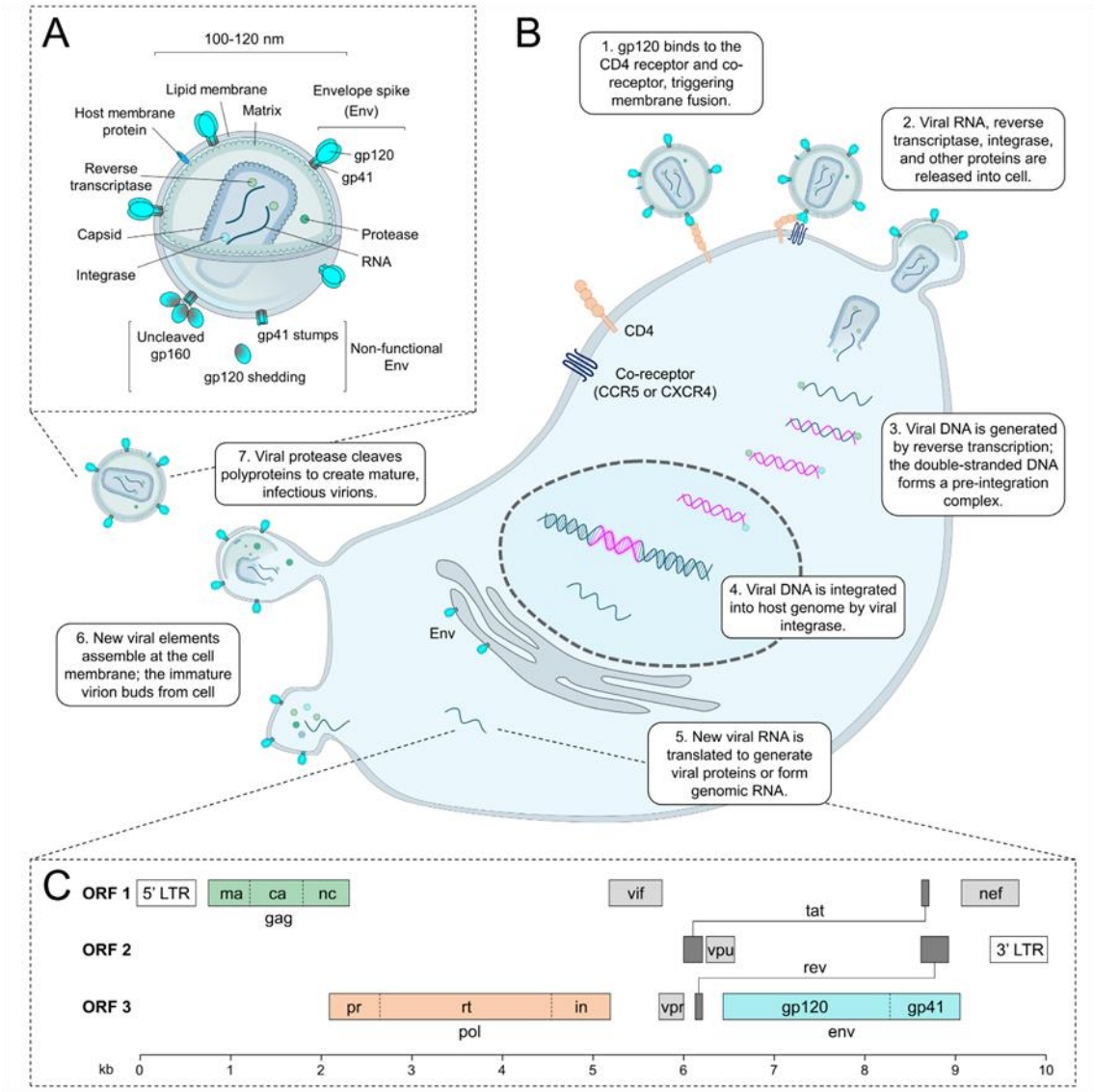


Figure 1-1: The HIV-1 structure and lifecycle.
(A) Structure of an HIV-1 virion. (B) The viral lifecycle of HIV-1 (C) The viral genome, approximately 9.8 kb, is transcribed over multiple open reading frames. MA, matrix; CA, capsid; NC, nucleocapsid; PR, protease; RT, reverse transcriptase; IN, integrase; Vif, viral infectivity factor; Nef, negative regulatory factor; Vpu, viral protein U; Vpr, viral protein R; Tat, trans-activator of transcription. Figure reproduced from the thesis of G. Seabright, University of Oxford [30].

1.2.3 **HIV infection and the host's immune response**

Frequent routes of transmission for HIV-1 infection are through the mucosal membranes in the vagina and rectum. Upon transmission, the virus spreads to the spleen, bone marrow and lymph nodes and in particular to the gut-associated lymphoid tissue (GALT) which has a high concentration of CD4+ T lymphocytes [31]. This migration and retention between the blood and the lymphoid tissue occurs via the interaction of Env with the gut-homing α4β7 receptor which has been shown to interact with the variable loops on gp120 [32]. This interaction triggers the production of the integrin LFA-1 which promotes the cell-cell transmission of HIV-1 through what is known as the virological synapse. The importance of this route of transmission is debated with studies involving a range of gp120 monomers and infectious virions not able to bind the α4β7 receptor [33]. However, the RV144 trial utilised gp120 monomers and it is thought that the patients that were able to raise a protective antibodies against the V2 region disrupted this interaction and disrupted HIV-1 trafficking [34]. Macrophages and dendritic cells can also be infected, it is thought this occurs in macrophages when a macrophage engulfs an infected T lymphocyte [35].

During the first 10 days of an HIV-1 infection the levels of viral RNA are virtually undetectable, this is known as the eclipse phase. After this initial phase the viral load increases rapidly as the virus spreads to the various parts of the body via dendritic-cell mediated dissemination, with viral loads peaking at day 21-28 of infection [31]. During this phase the levels of CD4+ T cells decrease as the virus is infecting and killing these cells, however the levels of CD8+ T cells increases. This triggers the production of antibodies via activation of B cells. This is known as seroconversion. As the adaptive immune system begins to respond to HIV-1 infection the viral load decreases and the CD4+ T cell levels rise, although in the GALT the CD4+ T cells do not rise in the same manner [31]. These antibodies are capable of controlling HIV-1 infection initially, however, the early antibody specificities are directed towards protein epitopes. The error-prone reverse transcriptase means that the viral population is rapidly mutating and therefore the selection pressure exerted by these early antibodies results in rapid escape mutations causing viral escape and the viral load to increase. If left untreated the CD4+ T cell count continues to decrease until a critical limit is reached [36]. After this point the immune system of the host is irreversibly compromised and opportunistic infections that are rare, and include virus induced cancers, such as Kaposi's Sarcoma, which result in death.

1.3 The COVID-19 pandemic and the life cycle of SARS-CoV-2.

1.3.1 The emergence of a novel coronavirus

The world health organization (WHO) is constantly monitoring viral animal reservoirs for potential viruses that could cross over to humans and have pandemic potential. Coronaviruses have existed in the human population for many years, and typically cause diseases such as the common cold [37]. In addition to mild infection, coronaviruses have pandemic potential. In the 21st century, the first coronavirus to emerge with pandemic potential was the virus termed severe acute respiratory coronavirus-1 (SARS-CoV-1) which appeared in Asia in 2003 [38]. Following this in 2012 Middle Eastern Respiratory Syndrome coronavirus (MERS-CoV) was shown to have crossed from camels to humans and additionally had pandemic potential [39]. Both of these outbreaks were relatively contained, however the WHO categorized coronaviruses as one of the circulating viruses in animals which had pandemic potential. In late 2019 a cluster of patients were identified in Wuhan, a city in Hubei province, China, that displayed similar symptoms to previous coronavirus outbreaks. A new virus was identified from airway epithelial cells of infected patients on January 12th 2020 [40] and the WHO declared a global pandemic in March 2020 [24].

1.3.2 The SARS-CoV-2 viral genome and its life cycle

The SARS-CoV-2 genome consists of single-stranded positive-sense RNA, which is around 7,000 nucleotides long [40]. The genome sequence was published rapidly and made publicly available, which enabled the rapid design of vaccine candidates [41]. The SARS-CoV-2 genome consists of 12 functional open reading frames (ORFs) with 11 protein-coding genes. There are many non-structural proteins, which are key in viral replication, these are summarized in **Appendix A** [42]**.** In addition, the SARS-CoV-2 genome encodes several structural proteins, which together form an infectious virion. These include; the Spike glycoprotein (S), which is responsible for angiotensin converting enzymes 2 (ACE2) binding and initiating cell entry [43], the Envelope protein (E) which aid in the assembly and release of virions [44], the Membrane protein (M), a transmembrane protein, and the Nucleoprotein (N), which is a key component for RNA packaging [45] (**Figure 1-2**).

When an infectious virion encounters a cell expressing ACE2 on its cell surface the S protein binds, with the host serine protease TMPRSS2 acting as a co-receptor [46]. The virus can then enter the cell either through direct fusion with the plasma membrane or by endosomal uptake and uncoating [47]. After the viral RNA is released into the host cell, the virus begins replicating its genome. First, polyproteins pp1a and pp1ab are translated,

which are cleaved by the Papain-like protease and 3C-like protease to form functional non-structural proteins such as Helicase and the RNA replicase–transcriptase complex (RdRp) [48]. RdRp is responsible for replication of structural protein RNA. Unlike HIV-1, the mutation rate of SARS-CoV-2 is significantly lower [19]. This can be attributed to the existence of proof reading mechanisms in the RNA polymerase, containing 3'-to-5' exonuclease activity, which removes mis-incorporated nucleotides [49]. In addition, the period of time in which SARS-CoV-2 can be transmitted is brief [50], so there is little time for viral evolution to occur in a host prior to transmission, whereas HIV-1 can undergo years of evolution and still transmit. The structural proteins are then translated, with S, E and M proteins being translated by ribosomes that are bound to the endoplasmic reticulum (ER). These proteins then enter the ER and undergo post-translational modification as would host cell-surface glycoproteins. These are displayed on the outside of the virion. The N protein is translated in the cytoplasm remains in the cytoplasm during viral assembly. The virion then buds into the ER Golgi intermediate complex (ERGIC) and is secreted via exocytosis [51].

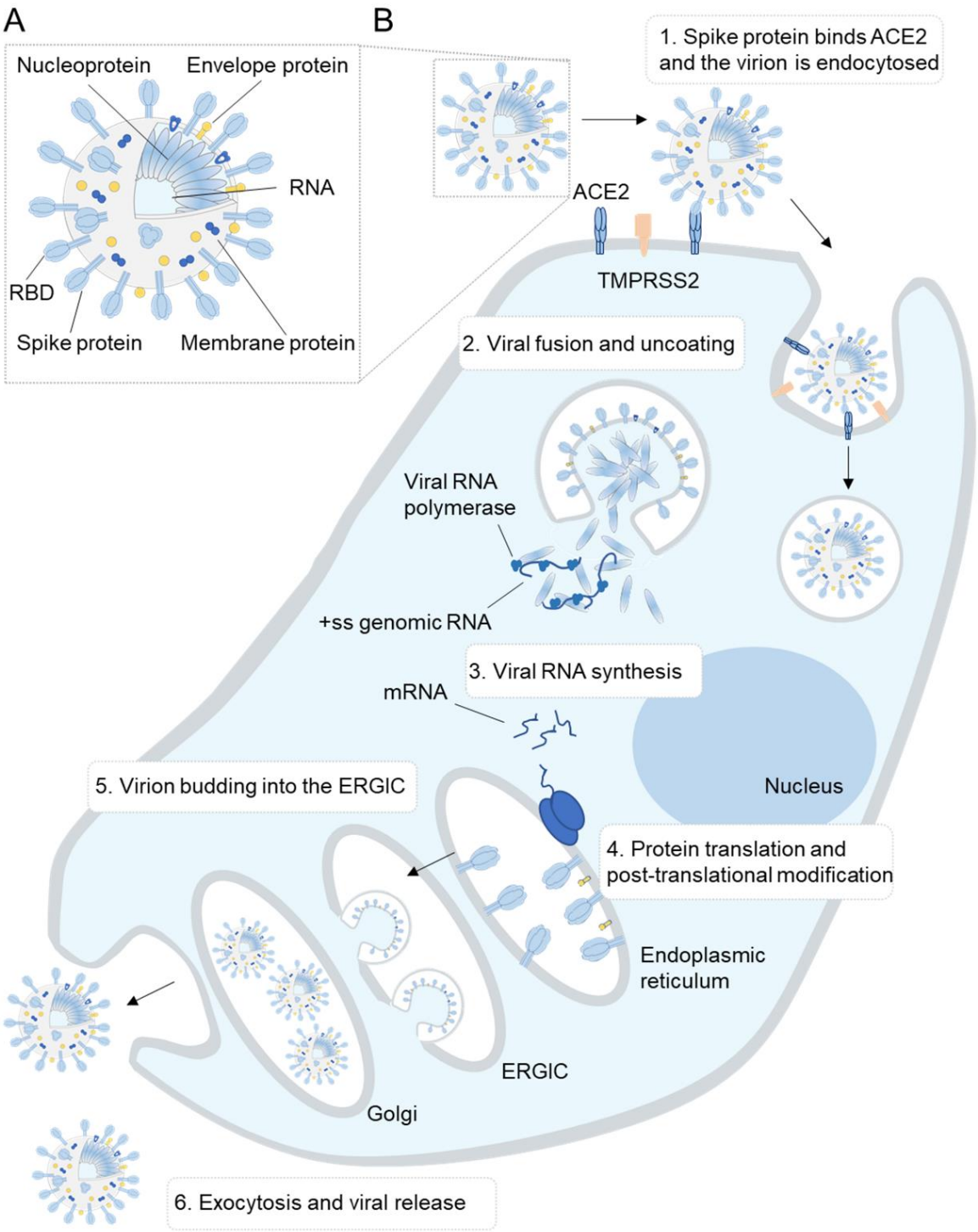


Figure 1-2: Life cycle of the SARS-CoV-2 virion.
(A) Cartoon displaying the contents of an infectious SARS-CoV-2 virion. (B) Viral life cycle of SARS-CoV-2. Key stages in the life cycle are numbered in the order which they occur.

1.3.3 SARS-CoV-2 pathogenesis and the host response to infection

Infectious SARS-CoV-2 particles are spread by either droplets, or potentially as an aerosol [52]. Once inhaled, the virus infects ACE2+ cells in the naso/oral cavity [53]. This phase of infection is usually asymptomatic as the virus is replicating in the upper respiratory tract, which is relatively shielded from the innate immune system [54]. Within 2 days to 2 weeks of initial infection COVID-19 symptoms begin to appear. These include an extensive range of different pathologies including fever, dry cough, shortness of breath, joint pain, and tiredness. As the virus infects the lower respiratory tract, via the airways, a strong innate immune response occurs. This includes a strong pro-inflammatory response as well as other complications, including pulmonary edema, Acute Respiratory Distress Syndrome (ARDS), different organ failures, and death [55].

Due to its high infectivity, SARS-CoV-2 has caused millions of deaths worldwide, however, the majority of people who become infected with SARS-CoV-2 are able to make a full recovery [56]. This means that the immune system is capable of clearing SARS-CoV-2 infection. This is achieved through both the innate and adaptive immune response. One of the first immune responses initiated against SARS-CoV-2 infection is via pathogen recognition receptors, such as Toll-like receptors, and results in upregulated interferon expression [54]. The humoral response against SARS-CoV-2 has been found to be similar to that against other coronavirus infections, involving IgM and IgG [57]. At the onset of SARS-CoV infection, B cells elicit an early response against the N protein, while antibodies against S protein could be detected after 4–8 days following infection [56]. These neutralizing antibodies targeting the S protein are extremely potent, and likely provide strong protection against subsequent infections. The role of T cells in SARS-CoV-2 immunity has also been demonstrated, and it is a combination of both B and T cell mediated immunity that likely provides an effective defence against SARS-CoV-2 [58].

Stimulating the immune system to respond as if encountering a real pathogen is the goal of vaccination efforts, therefore molecular mimicry of viral components is a key part of this process. As such, a deep understanding of the molecular make-up of the structural proteins of virions is important for the successful development of vaccines. Spike proteins have emerged as key vaccine targets, as during natural infection for both HIV-1 and SARS-CoV-2 potent neutralizing antibodies are raised, which, if induced prior to infection, can protect against disease.

1.4 Structure of Envelope glycoproteins

1.4.1 Class I viral fusion proteins and similarities between Env and S protein.

As viral spike proteins are critical for host cell invasion, and are exposed on the surface of the virion, they are frequently the target of neutralizing antibodies, which bind and inhibit the function of the spike protein, either by blocking the interaction with its desired receptor or by preventing the conformational changes required for membrane fusion [59,60]. As such, viral spike mimetics are key targets for vaccine design, with both SARS-CoV-2 and HIV-1 immunogens being based around S protein and Env, respectively. As these spike proteins are important antigenic targets, understanding the molecular details of spike proteins is critical in vaccine design. Whilst HIV-1 and SARS-CoV-2 are distinct in both the makeup of their genomes and in the disease that they cause, there are notable similarities in the structure of the surface spike glycoproteins, and advances to enable the study of HIV-1 Env enabled the rapid resolution of the SARS-CoV-2 S protein structure, which further aided the rapid development of successful SARS-CoV-2 vaccines.

A multitude of viruses use spike proteins to bypass the plasma membrane of a cell to insert their viral genomes into the host cell [61]. Different spike proteins are classified into four groups (I, II, III and IV), although as with most processes related to viruses there are likely many more mechanisms at work in nature. Class I fusion proteins are common in pandemic viruses and include HIV-1 Env, Influenza Haemagglutinin (HA), Lassa virus glycoprotein complex (GPC) and coronavirus spike proteins (**Figure 1-3**) [61,62]. Class I fusion proteins are defined by the mechanism by which they are able to fuse with the membrane. Following membrane binding, class I fusion proteins undergo conformational change, which exposes α-helical coiled-coils on the membrane fusion component of the spike protein. As discussed below this includes gp41 for Env, and S2 for S protein [63]. Examples of class II fusion proteins include flaviviruses such as Dengue and Zika virus, and Class III include herpesviruses such as HSV-1 glycoprotein B [64].

The commonality in the spike proteins of HIV-1 and SARS-CoV-2 has meant that innovations in HIV-1 Env research have aided in S protein structural determination. The key developments in this area are focused on the generation of soluble variants of spike proteins. Part of the function of spike proteins is to undergo conformational change, and as such they are not particularly stable proteins on the surface of virions [65]. To define the structure of viral spike proteins it is often beneficial to generate soluble versions of spike proteins. In addition, membrane bound vaccine candidates are often not translatable and classically, solubilized versions of viral spike proteins are required. Simply truncating the

transmembrane region of spike glycoproteins does not produced native-like proteins [66], and several stabilizing mutations are required. This work took many years for HIV-1 Env and proved essential for the resolution of the trimeric Env structure [67]. In addition, the lack of native-like architecture before these stabilizing mutations were applied likely played a role in the poor efficacy of early vaccine candidates against HIV-1 [68]. The researchers who solved the structure of the SARS-CoV-2 S protein for the first time had backgrounds in HIV-1 Env research and by applying similar strategies used to stabilise HIV-1 Env, were able to resolve the SARS-CoV-2 S protein structure within months of the publication of the S protein sequence [69].

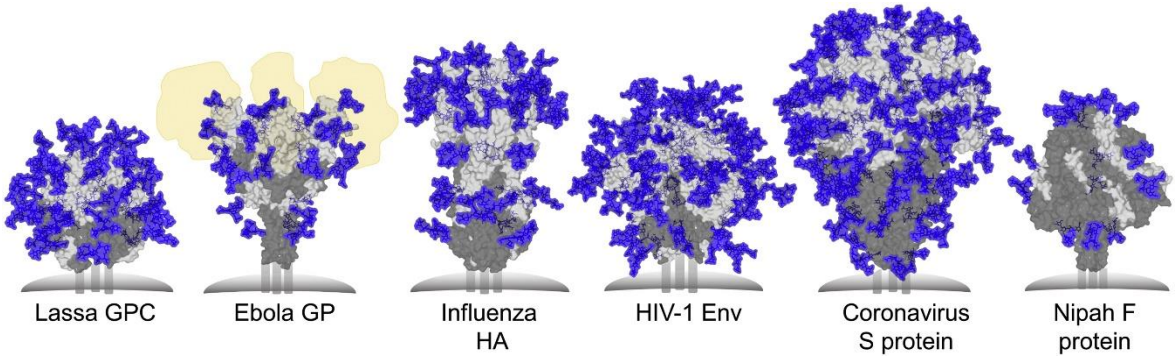


Figure 1-3: Examples of viral fusion proteins.
Left to right: Glycan shield models of Lassa virus GPC (PDB ID: 5VK2), Ebola GP (PDB ID: 5JQ3), A/H3N2/361/Victoria/2011 H3N2 Influenza virus hemagglutinin (PDB ID: 4O5N), BG505 SOSIP.664 HIV-1 Env (PDB ID: 4ZMJ), human coronavirus-NL63 (HCoV-NL63) S protein (PDB ID: 5SZS), Nipah F protein (PDB ID: 5EVM). Glycans and proteins are shown in blue and grey, respectively. The fusion protein subunit is shown in dark grey. The positions of mucin-like domains of Ebola GP are shown in yellow. Most predominant glycan compositions were modelled onto each N-linked glycan site, using pre-existing GlcNAc residues, where available, with $Man_5GlcNAc_2$ modelled on if compositional information was lacking. Figure reproduced from Watanabe *et al.* 2019 [61].

1.4.2 **HIV-1 Env and BG505 SOSIP.664**

The immense variation of the genome of HIV presents a considerable obstacle to the development of an effective, sterilizing vaccine. The failure of classical vaccination strategies has driven advances in our understanding of the molecular basis of viral infectivity and the host immune response. Lines of enquiry have included investigating correlates of protection arising from the RV144 trial such as the role of non-neutralizing antibodies [70–72]. These strategies have proven ineffective of protecting against HIV-1 infection [73], and as such, the focus of vaccination strategies has shifted towards developing immunogens capable of eliciting a protective broadly neutralizing antibody (bnAb) response [74].

Innovations towards the development of bnAb-eliciting immunogens have included the production of native-like trimers that mimic the envelope glycoprotein (Env) spikes expressed on the surface of the virion [75–78]. These soluble glycoproteins have revolutionized our understanding of the viral glycoprotein structure and have stimulated the design of new

immunogens. Encouragingly, these mimetics bind to a growing number of neutralizing antibodies isolated from infected patients, and they are being investigated as a platform for the next generation of immunogens. Armed with the detailed structure of envelope spike mimetics, detailed knowledge of the intricate network of post-translational modifications has been revealed [76,79–84].

Env is the sole viral protein expressed on the surface of the HIV virion and antibodies capable of binding the functional form of the viral spike can effectively neutralize viral infectivity [85–91]. As a class I viral fusion glycoprotein, Env consists of a trimer of heterodimers comprising gp120, which consists of five constant domains interspersed with hypervariable loops (**Figure 1-4A**), responsible for receptor recognition, and the transmembrane fusion glycoprotein, gp41. Env enables viral particle recognition of CD4+ T-cells, binding to the chemokine co-receptor, and ultimately fusion of the viral membrane with that of the target cell. To generate infectious viral particles, the viral spike must be extensively post-translationally modified by the host cell. These modifications include the formation of disulphide bonds (**Figure 1-4B**), extensive N-linked glycosylation (**Figure 1-4C** and **D**) and two proteolytic cleavage events, one after the signal peptide and the other between gp120 and gp41 (**Figure 1-4E**) [66].

Successful attempts to recombinantly mimic the Env-based epitopes of broadly neutralizing antibodies have involved solubilizing and stabilizing glycoprotein trimers utilising post-translational modifications and amino acid substitutions. Crucially, native-like Env trimers have been generated by the addition of a disulphide bond between gp120 and gp41 subunits and an isoleucine to proline mutation in gp41 which stabilise the prefusion conformation of Env. These immunogens are termed "SOSIPs" [76,92,93]. This approach was first successfully applied to the Env from BG505, a clade A/E transmitter/founder virus which was isolated in Kenya [76]. Sanders *et al.* additionally introduced the N332 site lacking from the BG505 isolate, as this site forms the target of a number of broadly neutralizing antibodies [76]. Whilst the BG505 strain is amendable to the modifications outlined above, newer iterations based on the above modifications has enabled the stabilization of many Envs from multiple clades [80].

The quaternary structure of natively folded viral spikes requires the correct formation of disulphide bonds in the mammalian secretory pathway. Slow folding events allow time for the extensive network of disulphide bonds to form and to "shuffle" using protein disulphide isomerase (PDI) to ensure the optimal disulphide bond configuration is achieved [94–96]. The correct pairing of disulphides is an essential requirement for infectious Env, as by knocking out PDIs *in vitro* the resultant protein is able to bind to CD4 but not undergo the

conformational changes required for membrane fusion [97]. The incorrect formation of disulphides is also promoted when viral spike mimetics remain uncleaved by furin, this not only impacts the resultant quaternary structure of the viral spike it also perturbs the glycan shield and therefore Ab binding [98].

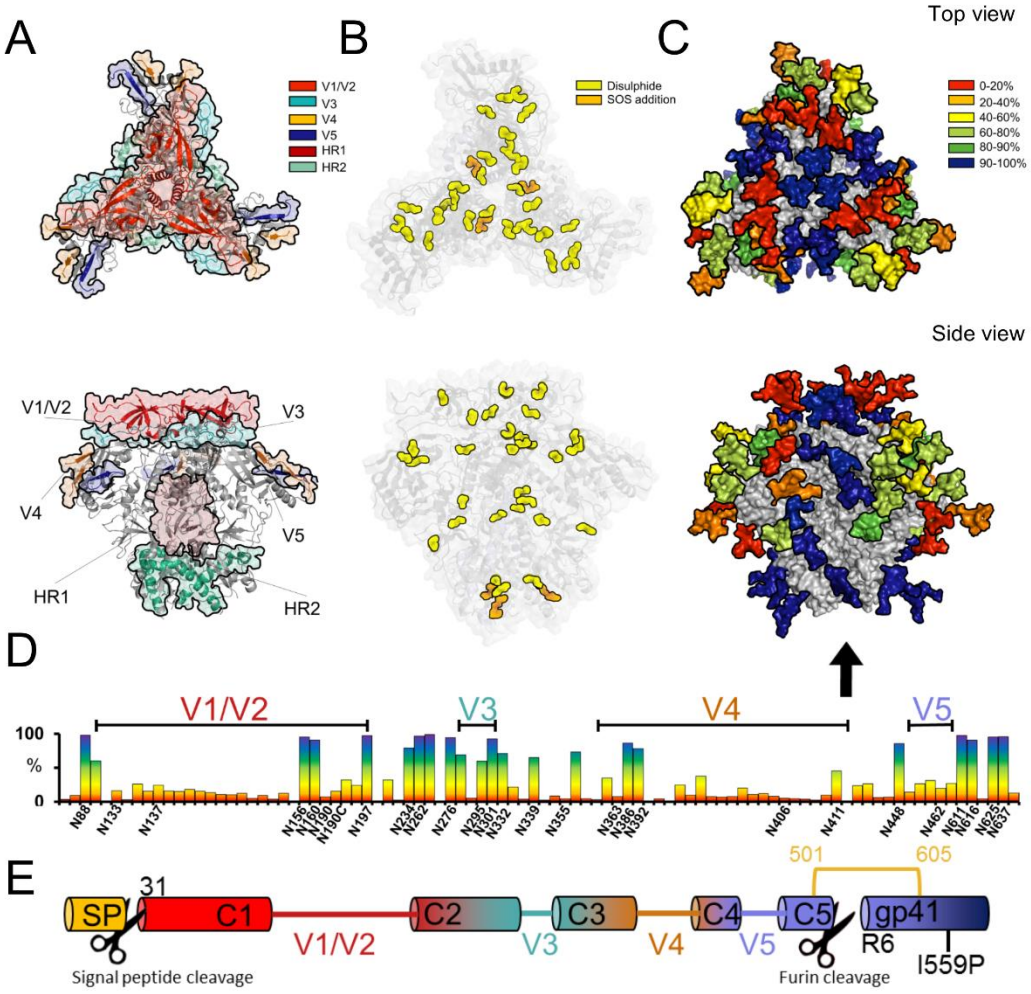


Figure 1-4: Representation of the post-translational modifications of BG505 clade A/E envelope glycoprotein.
Models were generated using the cryo-EM structure of BG505 SOSIP.664 PDB ID:5ACO [81]. (A) 3D representation of the variable loops on gp120 and the heptad repeats of gp41. (B) Canonical disulphide bonds found in BG505 [98] with the additional stabilizing disulphide bond found in BG505 SOSIP.664 shown in orange. (C) Conservation map of the glycans of BG505 SOSIP.664. The glycans were coloured according to their conservation across 4000 Env strains Huang *et al.* [99]. BG505 crystal structure with N-linked glycans modelled by Behrens *et al.* [100] (D) The frequency of potential N-linked glycan sites across the Env sequence with the PNGs found in BG505 SOSIP.664 labelled on the X axis. (E) Schematic showing the locations of proteolytic cleavage for the signal peptide and the furin cleavage site. Figure also published in Allen. *et al.* 2018 [29]

Another critical post-translational modification of Env is furin cleavage of the gp160 pro-protein. Furin cleavage is thought to occur in the trans Golgi network and is an essential step in the formation of functional correctly folded trimers [101]. Furin cleaves between the gp120 and gp41 subunits of the gp160 polypeptide, recognising the amino acid motif Arg-X-(Arg/Lys)-Arg [102]. Negative stain electron microscopy (NS-EM) of uncleaved trimers, which have been explored as viral spike mimetics, revealed that they do not naturally adopt

the classical trimeric architecture [103,104]. The perturbation of the structural integrity resulting from a lack of furin cleavage also has implications for the post-translational modifications of candidate immunogens. Analysis of N-linked glycosylation on a global and a site-specific level on a number of uncleaved structures described an elevation in glycan processing, most likely resulting from aberrant trimerization [79,105]. As broadly neutralizing antibody epitopes frequently contain N-linked glycans the aberrant glycosylation resulting from a lack of furin cleavage will have knock-on effects on the antigenicity of those trimers. Furthermore, the destabilization of the quaternary structure ablates the binding of quaternary-specific broadly neutralising antibodies (bnAbs) such as PGT151 and also reduces the affinity of other quaternary specific bnAbs, such as PGT145 [106,107].

The requirement for furin cleavage for correct assembly is an important feature in the expression of BG505 SOSIP.664 as the low levels of endogenous furin result in large populations of uncleaved trimers [108]. Binley *et al.* circumvented this problem by co-transfecting plasmid containing the *furin* gene concomitantly with Env [108]. In addition, it is important to optimise the protease cleavage step as much as possible to avoid the generation of uncleaved Env. The majority of HIV strains contain a furin cleavage site consisting of a REKR motif. A panel of mutations introduced to this region demonstrated that this is not the optimal motif for furin mediated Env glycoprotein cleavage. By replacing the REKR motif with 6 arginine residues the proteolytic separation of gp120 and gp41 is greatly improved [108]. By modifying the amino acid sequence in this way, it is possible to manipulate post-translational protease cleavage to allow for a larger amount of native-like material. Impressively, efficient furin cleavage has also now been achieved in production of gram quantities of clinical grade BG505 SOSIP.664 using a stable CHO cell line containing the target SOSIP, possessing an optimised furin cleavage site, and also the gene encoding furin [109].

As well using recombinant glycoproteins as immunogens, a further strategy currently under investigation in HIV-1 vaccine design is to use nucleotide-based approaches. For nucleotide-based vaccines, the *env* gene is administered and the trimers are expressed by the host. In order to utilise this approach, it is favourable to bypass the furin cleavage stage as it removes the requirement of co-transfection with furin. Although previously defined uncleaved trimers have produced large populations of misfolded trimers, several constructs now exist that are able to form native-like trimers without the need for furin cleavage. By replacing the furin cleavage site with a flexible linker it is possible to generate native-like trimers. With additional stabilizing mutations, native flexibly linked (NFL) trimer display native-like bnAb binding profiles and analogous glycosylation to BG505 SOSIP.664 [77,110,111].

1.4.3 **SARS-CoV-2 S protein and the 2P mutations**

Following the outbreak of the COVID-19 pandemic the scientific community rapidly refocused towards understanding this novel pathogen. As a result of a concerted effort from researchers across the globe, several vaccines have now been developed, with over 800 million first doses administered as of June 2021 [112]. This remarkable progress was possible due to years of prior research which was rapidly deployed. Several vaccines, including from Moderna and Pfizer/BioNtech, use stabilized spike protein sequences as the key immunogen. The key modifications of the S protein were first employed to resolve the structure of SARS-CoV-1 and MERS-CoV in 2017 [113], from Jason McLellan and Andrew Ward, and include several of the modification strategies outlined above; proline substitutions to stabilize the pre-fusion spike conformation, beneficial for neutralizing antibody binding [4,113], and the replacement of the furin cleavage site. These stabilizing mutations were employed by several labs to obtain a detailed structure of the SARS-CoV-2 S protein [69,114].

SARS-CoV-2 utilizes an extensively glycosylated spike (S) protein that protrudes from the viral surface to bind to angiotensin-converting enzyme 2 (ACE2) to mediate host-cell entry [43]. The S protein is trimeric class I fusion protein, composed of two functional subunits, responsible for receptor binding (S1 subunit) and membrane fusion (S2 subunit) [69,114]. The SARS-CoV-2 S protein is translated a single polypeptide and is cleaved at the S1/S2 interface by furin [115]. The S1 subunit is separated into the N-terminal domain (NTD) and the receptor binding domain (RBD), which is responsible for ACE2 binding (**Figure 1-5A**). Remarkably, the surface of the envelope spike is dominated by host-derived glycans with each trimer displaying 66 N-linked glycosylation sites. Stabilization of the trimeric prefusion structure was achieved using the “2P” stabilizing mutations [113] at residues 986 and 987, a “GSAS” substitution at the furin cleavage site (residues 682–685), and a C-terminal trimerization motif. This enabled the resolution of the 3D structure (**Figure 1-5B**) [69]. Further proline substitution at key regions of the spike protein further stabilized the pre-fusion conformation of the S protein, this was termed “Hexapro” as a result of the 6 proline substitutions and demonstrates the usefulness of proline substitutions in stabilizing viral spike proteins [116].

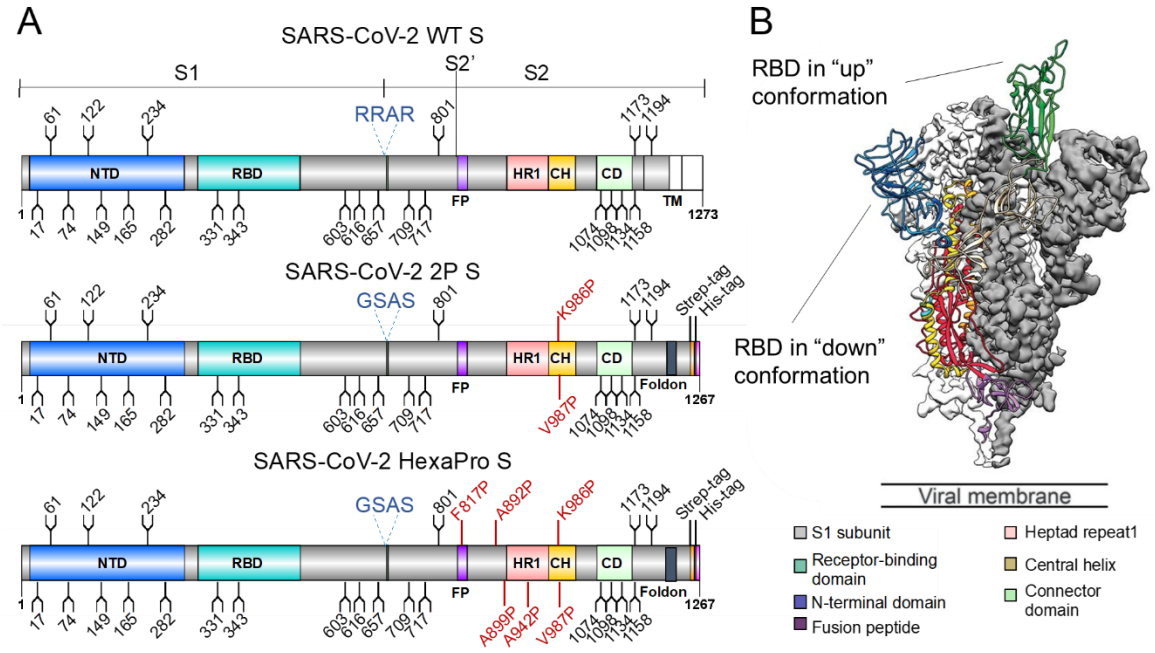


Figure 1-5: Schematic detailing the domains of the SARS-CoV-2 S protein and the 3D structure of the S protein.
(A) Schematic of the SARS-CoV-2 S protein illustrating the different domains and molecular features of the Spike protein. N-linked glycans are labelled as black sticks. The top schematic represents the WT S protein found on viruses and below are two stabilized versions, with proline substitutions labelled red. Schematics produced in collaboration with Himanshi Chawla, University of Southampton. (B) Structure of the SARS-CoV-2 S protein obtained using the 2P version of the S protein, reproduced from Wrapp *et al.* 2020 [69].

Whilst the 3D structure and stabilization of viral spike glycoproteins is critical for vaccine design, there are features of spike proteins that are difficult to visualize using structural techniques such as cryo-EM and X-ray crystallography. As viruses are obligate parasites, they hijack their host's secretory machinery to attach post-translational modifications. Both HIV-1 Env and SARS-CoV-2 S protein are covered in N-linked glycans, therefore it is not possible to fully define the structure of viral spike glycoproteins without also understanding how they are glycosylated.

1.5 Glycosylation of the host cell and the glycan shields of viruses

Glycosylation is a fundamental post-translational modification which can often be overlooked due to its complexity. The process of glycosylation is a broad term that refers to the modification of a biological molecule with one or more monosaccharides, which are referred to as glycans, and a protein with a glycan attached is referred to as a glycoprotein.

Many host proteins are modified with glycans, especially cell surface and secretory glycoproteins, and glycans play a multitude of roles including in protein folding and stability, and can even govern the effector functions of immunoglobulins [117]. Viral glycoproteins co-opt glycosylation to stabilize the protein fold, but also employ glycosylation for other roles beyond those of host glycoproteins [61]. As glycans are host derived they are often poorly immunogenic, and their presence can be used to shield the highly antigenic protein surface [118]. This likely explains the high abundance of N-linked glycans across all the viral glycoproteins outlined above [61]. To design vaccine candidates that are heavily glycosylated it is important to understand how the vaccine antigens are glycosylated, how the glycosylation compares to that of glycoproteins on the infectious virion, and if the two can be reconciled. This process requires understanding how the host glycosylates its own protein, as viruses hijack this process, and then applying bespoke techniques to define both immunogen and virus glycosylation.

1.5.1 **Glycans and their building blocks**

A monosaccharide is the smallest unit of a carbohydrate and are typically classified according to the number of carbon atoms they contain. When attached together as part of a glycan, monosaccharides typically adopt a cyclic structure, with the majority of mammalian N-linked glycans consisting of 5 atom rings, furanose, or 6 atom rings, pyranose (**Figure 1-6**). The precise arrangement of side groups defines the exact name of each monosaccharide, many common monosaccharides only differ in the orientation of hydroxyl groups and are therefore isomers of one another. The 6 carbon monosaccharides are often termed “hexose” (Hex) and include glucose, galactose, and mannose, with the only difference between glucose and mannose being a different orientation of the hydroxyl group located on the carbon at position 2 (**Figure 1-6A**). Different chemical groups can be attached to the central ring, for example the addition of an N-acetylamine at position 2 of a hexose, referred to as N-Acetylhexosamine (HexNAc). Additional modifications can add additional complexity to monosaccharides, for example sialic acid, or neuraminic acid (NeuAc) contains several extended side chains, one of which can confer a negative charge to the monosaccharide, which has implications for the function of a glycan to which it is attached [119–122].

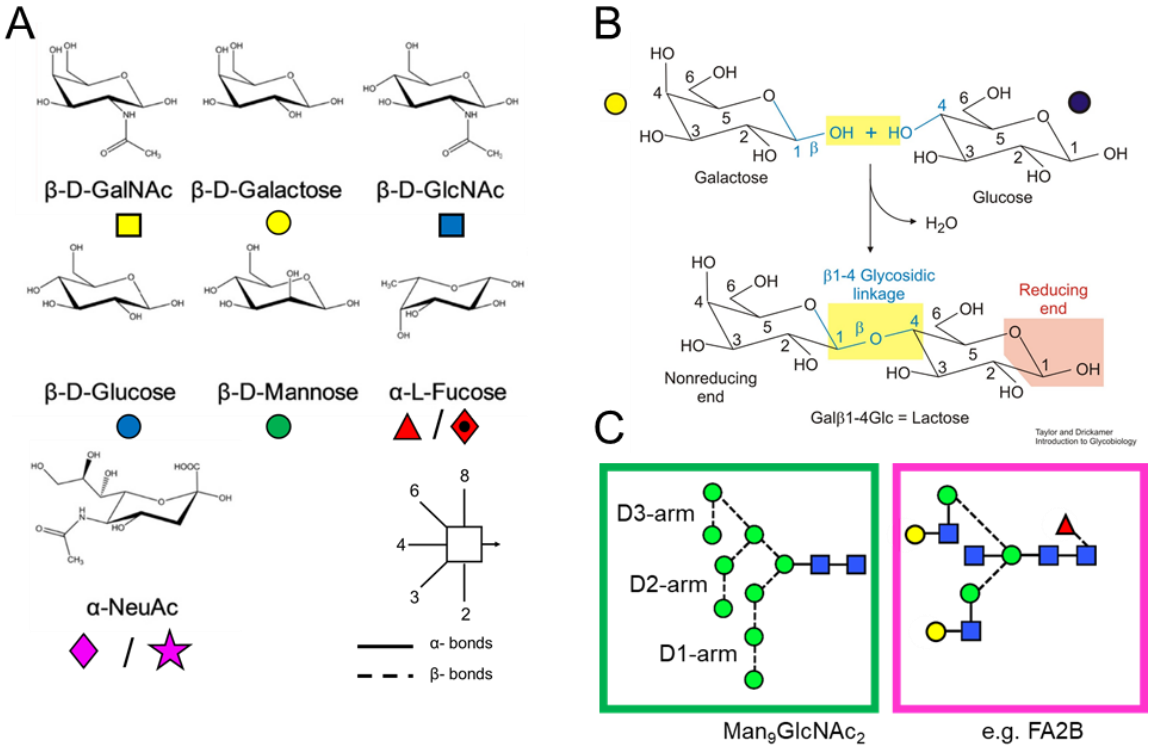


Figure 1-6: Monosaccharide structures, how they combine to form a glycan and commonly used conventions to simplify their presentation.
(A) Common monosaccharides that form N-linked glycans on human glycoproteins. The chemical structure is shown above its corresponding name. Displayed underneath is the symbol used to represent each monosaccharide, used for both the Symbol nomenclature for glycans (SNFG) [123] and the Oxford classifications. Where two symbols are shown the left corresponds to SNFG and the right is Oxford. (B) Formation of a glycosidic bond reproduced from Introduction to Glycobiology [124] to demonstrate its formation, and highlight how glycosidic bonds are defined, using the formation of lactose as an example. (C) Two example N-linked glycan structures shown using SNFG symbols and Oxford bond angles.

The linkages between monosaccharides are important as they often dictate enzyme specificity and lectin binding. An example of a glycosidic bond is shown in **Figure 1-6B.** Here, the bond is formed between the hydroxyl group attached at carbon 1 of the galactose (Gal) and position 4 of the glucose (Glc), and the bond forms when the C1 hydroxyl is in the β orientation. This bond is therefore referred to as a β1-4 glycosidic linkage, and the disaccharide shown is referred to as Galβ1-4Glc. Many monosaccharides can link together and form a glycan, and two examples of such structures are shown in **Figure 1-6B**. To avoid drawing the full chemical structure, it is common for glycobiologists to represent glycans as a series of connected shapes. This thesis uses a combination of two naming conventions, the Oxford and the SNFG. These two share many features, such as the shape and colouring of Hex and HexNAc residues, however they differ slightly as shown in **Figure 1-6A** [125,126]**.** Additionally, the Oxford system displays the angle of each linkage to reflect the glycosidic bond it represents [127].

1.5.2 The human N-linked glycosylation pathway

The two most salient mechanisms of glycosylation in immunogen design, against viruses, are those attached to asparagines, referred to as N-linked glycans, and those attached to serines and threonines, or O-linked glycans. N-linked glycans are more common on viral glycoproteins, with half the mass of the HIV-1 Env consisting of N-linked glycans [80]. The N-linked glycan processing pathway is dictated by a series of enzymes and is different from processes such as transcription and translation. Whilst there is an amino acid sequon that dictates the attachment of N-linked glycans, the processing of N-linked glycans is detached from the primary amino acid sequence. There are two main types of enzyme involved in glycan processing, broadly grouped into glycosyltransferases, which attach new monosaccharides, and glycosidases which cleave glycosidic bonds. The expression level of these enzymes dictates the precise processing state of an N-linked glycan, in addition to a range of other parameters, such as metabolite availability [128]. It is important to note here that a single glycan is not uniformly attached to a single glycan site. For example, the single potential N-linked glycan site located in the IgG Fc, in a single recombinant expression of protein, can have well over 10 different glycan compositions attached, despite containing an identical amino acid sequence and structure [129]. As glycosylation is dictated by the activity of glycosidases and glycosyltransferases, the producer cell has a profound impact on the glycans present on a glycoprotein [105]. Glycosylation pathways between different species diverge, however there are conserved features between species as diverse as insects and humans [130]. As the work described in this thesis focuses on human pathogens, the human N-linked glycan pathway will be discussed in detail, noting points of diversion when appropriate.

N-linked glycan synthesis occurs prior to attachment to the glycoprotein, in the cytoplasm, where a lipid, dolichol, is phosphorylated [131]. This acts as an acceptor for a series of monosaccharide additions, beginning with two N-acetylglucosamine residues, followed by 5 mannose residues. The $Man_5GlcNAc_2$ attached to dolichol phosphate is then flipped inside the ER lumen and an additional four mannose residues are added, followed by three glucose residues [132]. The translation of mRNA which encodes for proteins that will either be secreted or inserted into a membrane occurs on ribosomes associated with an ER-translocon. This means that as the nascent polypeptide chain is synthesized it enters the ER lumen. The glycosyltransferase for N-linked glycans is associated with the ribosome-translocon complex and as the polypeptide chain emerges into the ER lumen, the $Glc_3Man_9GlcNac_2$ glycan is transferred from the dolichol phosphate to asparagines on the polypeptide chain [133]. Not all asparagines act as acceptors for N-linked glycans, only those followed by a serine or threonine at the third position following the asparagine. The second position following the asparagine can consist of any amino acid except proline, and the N-

glycan sequon is defined as NxS/T [134]. Following attachment of the precursor N-glycan to the asparagine residue the protein begins to fold. The glucose residues are then sequentially removed and act as folding sensors. When the final glucose residue has been removed, this signals to the Calnexin/Calreticulin cycle that the protein is correctly folded [135].

Following protein folding, the glycan processing pathway continues with the trimming of mannose residues from $Man_9GlcNAc_2$ to $Man_8GlcNAc_2$ by ER α mannosidase I [136]. The protein then exits the ER and transits through the Golgi apparatus, where the remainder of glycan processing takes place. In the ER/Golgi intermediate compartment (ERGIC) and early Golgi further mannose trimming occurs through the action of a series of mannosidase enzymes that trim 8 mannose residues down to 5 [136]. Glycans ranging from $Man_9GlcNAc_2$ to $Man_5GlcNAc_2$ are considered as oligomannose-type glycans and their significance is discussed more below. From this point glycosyltransferase enzymes can interact with the glycan, the first of which is N-acetylglucosaminyltransferase I (GlcNAcT I or GnTI) [137]. This transfers an N-acetyl glucosamine residue onto the D1 arm of the $Man_5GlcNAc_2$ glycan (**Figure 1-7**). Up until this point, N-linked glycan processing follows a linear path, however from here the pathway diverges. Following this processing step, fucosyltransferase can act upon the $Man_5GlcNAc_3$. If no further mannosidase activity occurs on the D3 and D2 arms, this results in the generation of hybrid-type glycans, whereby the D3 arm can continue to be processed, whereas the D1 and D2 arms remain intact [138]. If Golgi α mannosidase IA-C can act to remove the D1 and D2 mannose residues, then further glycan processing is possible. Any glycan from this processing step onwards is referred to as a complex-type glycan. As the name suggests, from this point a diverse range of different glycan processing stages occur. At any point during these processing stages the protein can exit the Golgi apparatus, terminating glycan processing.

From the $Man_3GlcNAc_3$ glycan several different glycan processing steps can occur. Firstly, further GlcNAc transfer can occur. As additional GlcNAc residues are added, the glycan becomes more branched. The number of GlcNAc added to $Man_3GlcNAc_2$ define the number of glycan antenna, for example, common antibody Fc N-linked glycan compositions contain 4 GlcNAc residues, two make up the core and the remaining two are branching, so this glycan would be referred to as a biantennary glycan [139,140]. Following GlcNAc transfer, galactosyltransferases can act and add galactose to the GlcNAc, forming typically a β1-4 linkage [141]. This galactose can then act as an acceptor for sialic acid. In addition, the core GlcNAc residue, which is bonded to the protein, can be modified by fucosyltransferase [142]. As mentioned above, there is no defined exit point of the glycan processing pathway, and mature glycoproteins can contain a mixture of glycans of many different processing states,

so the same glycan site across multiple proteins could contain a glycan that is fucosylated or afucosylated, biantennary or triantennary and so on. This results in an extremely heterogeneous population of N-linked glycans on the mature glycoprotein. It is important to note that there are many more N-glycan modifications possible during ER and Golgi transit, but the modifications outlined above are the most common and relevant for the cell lines outlined in this thesis.

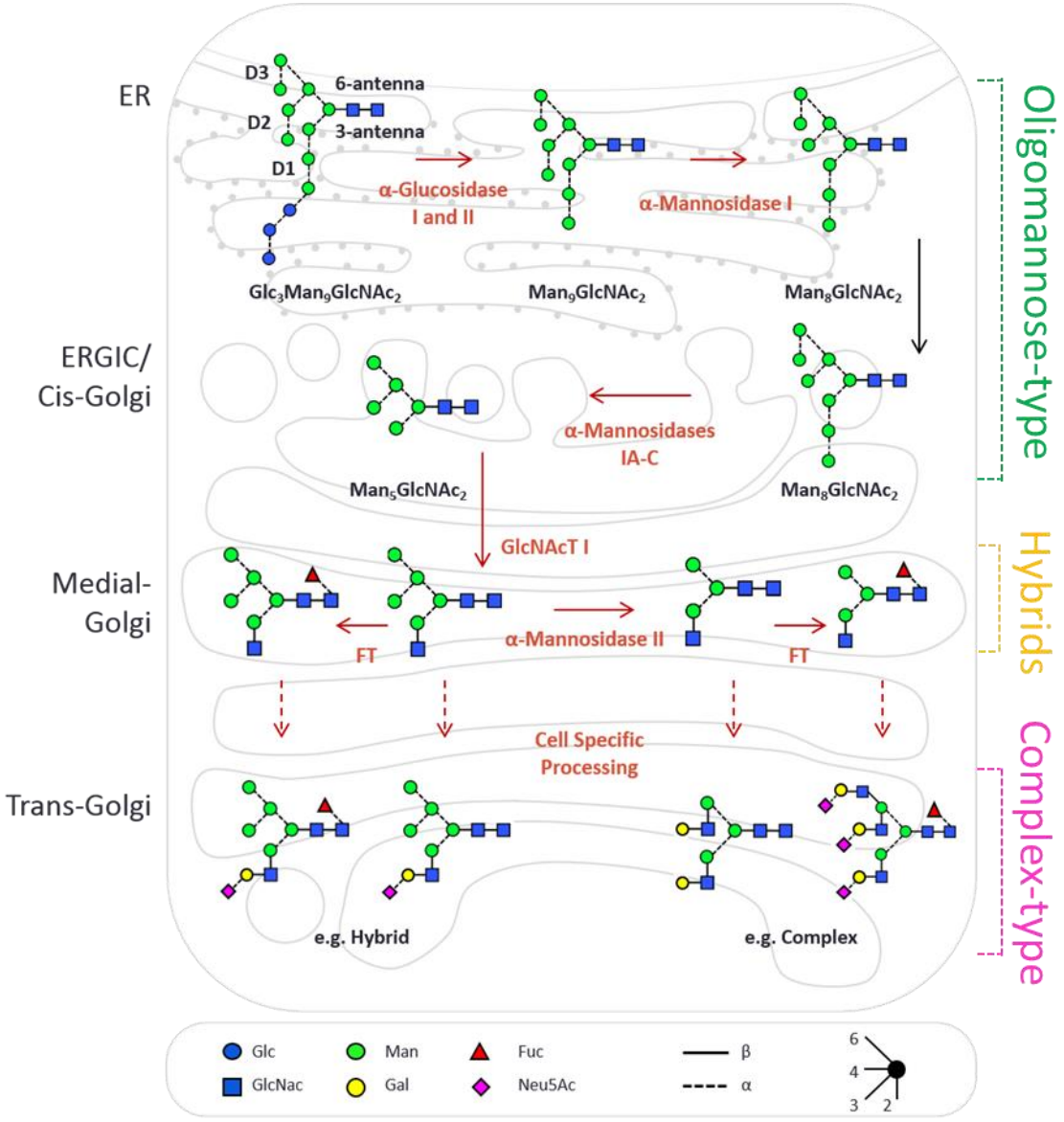


Figure 1-7: An overview of the mammalian N-linked glycan pathway.
Glycans are labelled according to the centre for glycomics defined shapes and the Oxford system linkages. Enzymes are labelled in red.

1.5.3 Viral N-linked glycosylation.

The heterogeneity of N-linked glycosylation resulting from the complex activity of glycosyltransferases means that N-linked glycan processing cannot be taken for granted during vaccine design. As different cell lines can have different glycosyltransferase expression levels, the resultant glycan distribution, also known as the glycoform, can differ from cell to cell on the same protein. As viral glycoproteins such as the SARS-CoV-2 S

protein and HIV-1 Env are extensively glycosylated, it is important to define the antigenic surface of viral spike proteins, with respect to both protein and glycan. Important considerations regarding glycosylation for immunogen design include the different producer cells of virus and immunogens, the occupancy of N-linked glycans, and if the occupancy differs between immunogen and virus. Finally, it is important to investigate differences between viral glycoproteins and healthy host cell proteins, which may potentially identify routes by which to prevent or treat viral infection.

1.6 Methods to analyse the glycosylation of candidate immunogens and targeted viruses

The amount of information pertaining to viral glycan shields has increased substantially with the development of powerful techniques such as in-line liquid chromatography-mass spectrometry (LC-MS) and high pressure high-performance liquid chromatography (HPLC) instruments. A key study in the characterisation of the HIV-1 Env glycan shield involved the application of a range of techniques to study the glycosylation of the BG505 SOSIP.664 immunogen candidate on both a global and site-specific level [100]. Work by Sanders *et al.* showed that this soluble cleaved immunogen has antibody binding properties analogous to viral material [76]. This enabled the generation of large quantities of soluble Env and catalysed the structural characterisation of the protein [67]. These techniques enabled high-resolution determinations of Env, however, the N-linked glycans were left unresolved due to their intrinsically flexible nature as well as their heterogeneous compositions. Behrens *et al.* characterised the glycosylation of this protein using a three-stage methodology that has formed the basis for subsequent glycan analysis of viral glycoproteins [100].

1.6.1 Quantitative glycan analysis of the entire glycome of a glycoprotein

The first stage involved the release of N-linked glycans from the protein using peptide N-glycosidase F (PNGaseF) [142]. This can be performed in-solution or by first subjecting the protein to SDS-PAGE and excising the gel band which corresponds to the glycoprotein of interest. The resultant released glycans are then fluorescently labelled using either 2 aminobenzoic acid (2-AA) or procainamide and then subjected to hydrophilic interaction liquid chromatography ultra-high-performance liquid chromatography (HILIC-UPLC) analysis. The procainamide or 2-AA label attaches to the reducing end of the glycan with a 1:1 stoichiometry, and as such provides a quantification of the total glycoforms of all sites across the desired protein (**Figure 1-8A**) [143]. Following initial analysis of a labelled

aliquot of released glycans a UPLC chromatogram is obtained, however, this information alone is not sufficient to identify which glycan corresponds to which peak.

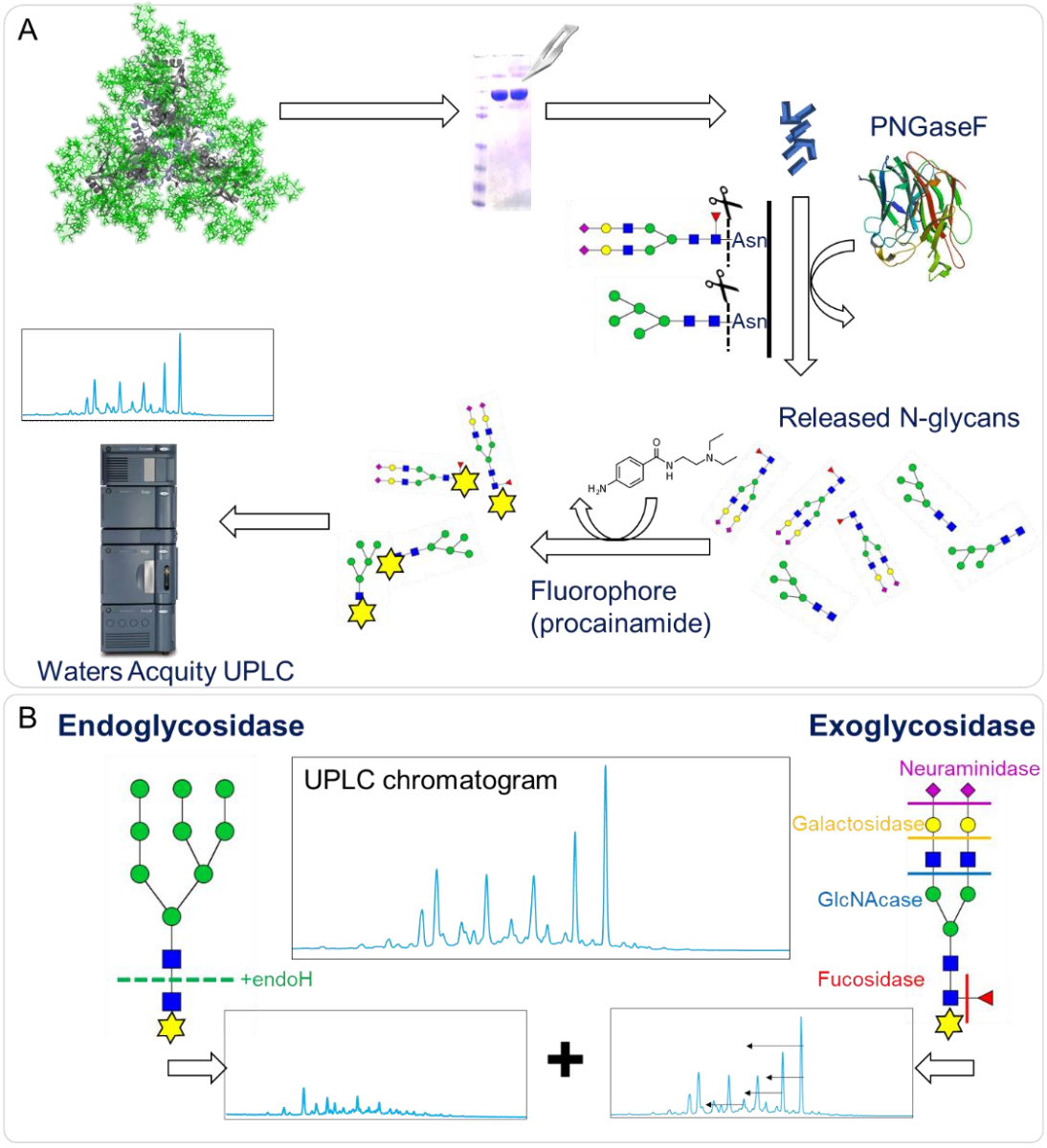


Figure 1-8: Diagram demonstrating outline of UPLC experiments.
(A) Schematic detailing sample preparation for UPLC, beginning with SDS-PAGE of target protein and subsequent labelling of the released glycans. (B) Strategies used for glycan identification of UPLC data using exoglycosidase and endoglycosidase enzymes. Glycans are presented according to the scheme outlined in Figure 1-6 using SNFG nomenclature.

Further incubation with glycosidases that cleave only specific glycosidic bonds can be used to determine which HPLC peaks correspond to certain glycans. The most commonly used for Env HPLC analysis is endoglycosidase H (endoH), which is specific for glycans that are mannosylated on the D1 and D2 arms [144], which are found only on oligomannose and hybrid-type glycans (**Figure 1-8B**). This enables the oligomannose content of the total glycan pool to be determined and has been applied to study a range of Env immunogen's oligomannose-type glycan contents [80,100,103,109,145,146]. In addition to

endoglycosidases, a range of exoglycosidases have been characterized, which cleave specific glycosidic bonds between specific monosaccharides. These have proven invaluable tools for studying N-linked glycan abundances by UPLC. Following analysis of the untreated released glycan pool, it is possible to determine the abundance of particular glycan modifications. One example is the quantification of sialic acids and the linkages that connect them to the remainder of the glycan. Different neuraminidases can be used which possess different specificities, for example a specificity towards α2,3 or α2,6 linked sialic acids [103]. Following exoglycosidase treatment, HPLC analysis is performed again and, unlike endoglycosidases where the procainamide label is cleaved from the N-linked and the signal is depleted, exoglycosidases reduce the size of the labelled glycan, which impacts their hydrophilicity, and causes peaks in the HPLC chromatogram to shift. By quantifying the peaks that shifted in the untreated sample it is possible to quantify particular monosaccharide modifications across the entire released glycan population for a particular protein.

1.6.2 **Glycan structure determination using mass spectrometry**

In addition to UPLC, other methods can be applied on enzymatically released glycans to understand the composition of the glycome of a glycoprotein. Mass spectrometry is a valuable tool to study glycosylation as it is possible to simultaneously acquire data from heterogeneous mixtures, assigning mass to individual glycans with high accuracy allowing the composition to be determined [147]. There are many different mass spectrometry approaches that have been successfully used to study glycans, each with their own advantages and disadvantages, and the following sections will detail the techniques used in the results chapters, acknowledging other approaches where necessary.

The preparation of released glycan aliquots is similar to the sample preparation for UPLC. The glycoprotein of interest is first subjected to SDS-PAGE and gel bands corresponding to the target of interest is excised and the N-linked glycans are removed by PNGaseF in-gel. For the released glycan analysis outlined in **Chapter 3,** methods developed by David Harvey were used to identify the glycan compositions of HIV-1 Env [148]. Prior to injecting released glycans into the mass spectrometer, ammonium phosphate, is added to the sample. One of the difficulties of glycan mass spectrometry is that glycans are poorly ionizable [149], and the generation of ions is a critical first step of this process. The addition of ammonium phosphate promotes the formation of phosphate adducts which ionize more readily. To generate ions, electrospray ionization (ESI) is commonly used as it enables the conversion of molecules in a solution into ions. In this approach, a narrow capillary needle is used, to which a high voltage applied. Under high voltage, an aerosol

known as the Taylor cone, emits a jet of liquid drops [150]. The solvent from the droplets progressively evaporates, leaving them more and more charged. As the charge of the droplets reaches a critical point, known as the Rayleigh limit, the droplet explosively dissociates, leaving a stream of charged ions [150]. These ions then enter the mass spectrometer.

The instrument used for these analyses was the Waters Synapt G2Si (**Appendix B**), which uses time-of-flight to establish the mass to charge (*m/z*) of ions. In this process ions are accelerated by an electrical field of known strength and, as such, ions with the same charge possess the same kinetic energy. The velocity of the ion depends upon both the mass and the charge of the ion, with ions of a heavier mass reaching the detector at a slower speed. The time taken for ions to reach the detector is proportional to the *m/z* of the ion [151]. In order to determine the mass of a particular ion and the composition of a glycan structure, the charge must be assigned. This can be done using the isotope distribution of a particular species. Biological molecules, including glycans, consist of many carbon atoms, and whilst the majority of carbon exists as carbon 12, larger isotopes are also present. As mass spectrometry can measure *m/z* with high accuracy, when a molecule is ionized it is possible to resolve the abundance of different isotopic populations of that molecule. The detected peak with the lowest *m/z* is referred to as the monoisotopic peak and represents every atom in the molecule consisting of the smallest isotope possible. In addition to this peak, biological molecules typically produce multiple isotope peaks when analysed by mass spectrometry. Each subsequent isotope peak represents an additional neutron. As mass to charge is measured during mass spectrometry, the isotope distributions enable the charge of the ion to be determined. The spacing between isotope peaks determines the charge state, so if the difference in *m/z* between isotope peaks is ~0.5 then the ion of interest has a charge (*z)* of two, as each additional neutron adds ~1 to the mass, so the *m/z* is ~1/2. By determining the charge state of a particular ion of interest it is possible to convert *m/z* into mass. This can then be used to determine the mass of that ion. For released glycan analysis the mass of all ions detected is first calculated, and these masses are cross-referenced to match any known N-linked glycan structures. An example of a released glycan dataset is shown in **Figure 1-9A.**

As many glycan structures are isomers, such as hexoses like mannose and galactose, often the composition alone is not sufficient for complete glycan identification. Whilst compositional analysis combined with an understanding of the particular glycan structures that make up the N-glycan processing pathway to infer certain structures is informative, mass spectrometry is also a powerful technique for the structural assignments of N-linked glycans. To define the structure of a glycan, ions are fragmented within the mass spectrometer. This fragmentation occurs in a predictable manner and the fragments can be

assembled to define the structure of an ion. For both released glycan analysis and site-specific glycan analysis (outlined below) a particular target *m/z* of interest is first isolated and then selectively fragmented. In the Synapt G2Si this is done using a quadrupole mass analyser. This consists of four cylindrical rods and is able to select ions by their *m/z* using oscillating electric fields, whereby only ions of a specified *m/z* are stable, and other ions are not able to transmit through the quadrupole [151]. With the target *m/z* isolated it is then possible to fragment the ion. For these experiments collision induced dissociation (CID) was used. In this approach, ions are accelerated to increase their kinetic energy and collide with gas molecules inside a chamber filled with an inert gas, such as nitrogen, helium, or argon. This causes fragment ions to be generated which can be used to assemble the structure of the ion isolated [152]. This process is also referred to MS2 or MSMS, with the initial detection of ions representing MS1. An example of how this approach can be used to resolve the structure of a particular glycan is shown in **Figure 1-9B** and **C** With this analysis a sample-specific glycan library can be generated which can be used to reduce the false-positive rate of the site-specific analysis [100].

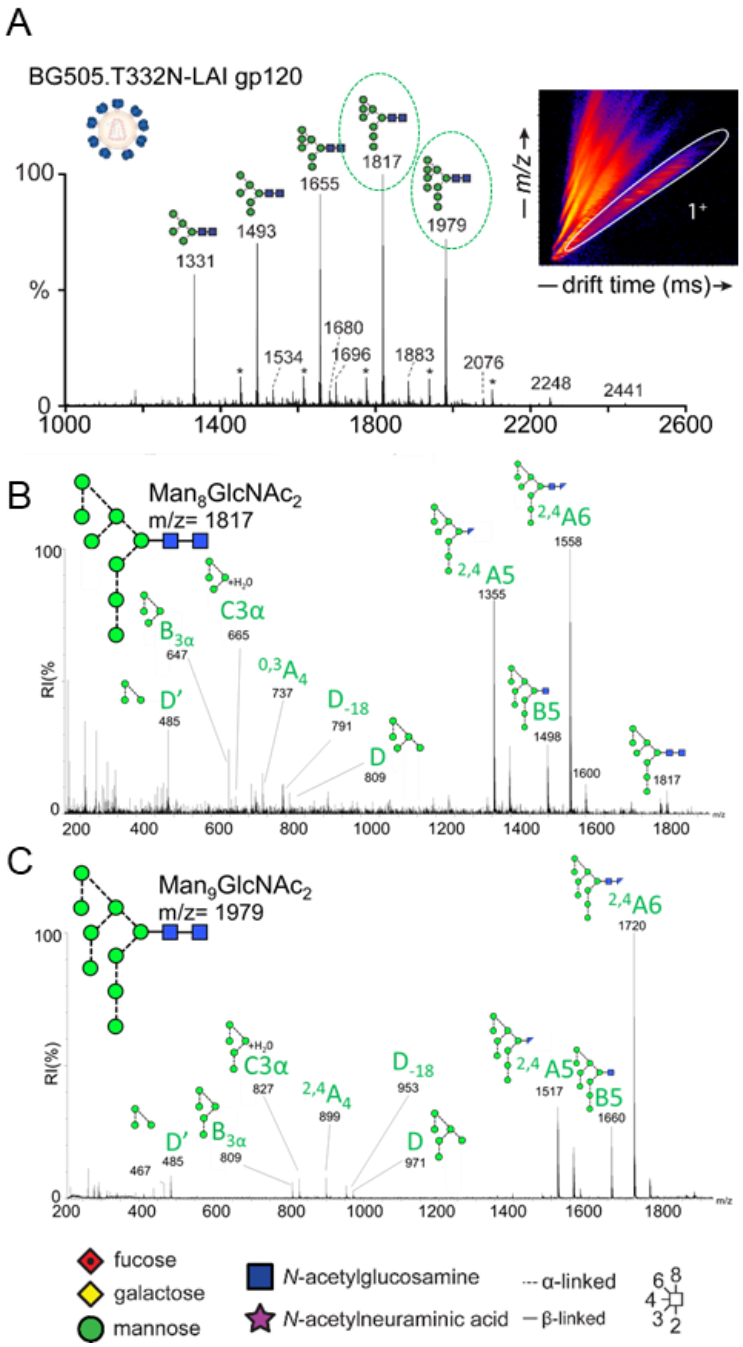


Figure 1-9: Ion mobility ESI fragmentation pattern for two oligomannose-type glycans.
(A) Example of a mass spectrum from a released glycan analysis experiment with oligomannose-type glycans labelled on the *m/z* corresponding to each structure. Adapted from Struwe, Chertova and Allen *et al.* 2018 [153] (B) MSMS fragmentation spectrum for $Man_8GlcNAc_2$. *m/z* peaks corresponding to fragments of the N-glycan are labelled with the appropriate glycan fragment. Cross-ring fragments are notated by a glycan dissected diagonally. (C) MSMS fragmentation spectrum for $Man_9GlcNAc_2$ labelled identically to (B).

1.6.3 **Glycopeptide mapping using liquid chromatography-mass spectrometry**

Whilst important information can be gained from understanding the total glycan populations using UPLC and released glycan mass spectrometry, to fully map the structure of a viral glycoprotein it is necessary to determine the distribution of glycans at each potential N-linked glycan site (PNGS). This requires both peptide mapping and glycan identification within a single experiment and techniques to determine the site-specific glycosylation of glycoproteins has advanced with the development of highly sensitive mass spectrometers and analytical software capable of handling complex datasets [154]. There are many different methods to perform site-specific glycan analysis, regarding sample preparation, sample analysis and downstream data analysis. The main methods used in this thesis were first developed by Behrens *et al.* in 2016 for analysis of BG505 SOSIP.664 and were further developed to investigate other parameters such as site-specific glycan occupancy [100,153,155]. The initial analysis performed by Behrens *et al.* used a mass spectrometer that measured *m/z* using time of flight, however all analyses performed in this thesis were obtained using an Orbitrap Fusion Tribrid mass spectrometer connected to an EasySpray nano-LC. To analyse the glycan distribution at a site-specific level it is necessary to first isolate each N-linked glycan site from one another. This is achieved by taking an aliquot of protein and digesting it with proteases with particular specificities. Prior to digestion, the sample is first unfolded, using a denaturant such as urea, and also reduced and alkylated to break, and prevent the reformation of, disulphide bonds. This process allows for optimal proteolytic cleavage. An example of an enzyme used in such experiments is trypsin, which has a specificity for lysine and arginine. Using proteases with particular specificity prevents complete degradation of the protein and results in a pool of peptides that can be detected by mass spectrometry [156,157]. Broadly, the goal of the sample preparation is to generate a peptide containing only a single PNGS to allow for the glycan distribution at each different PNGS to be determined.

Site-specific glycan analysis involves the generation of complex datasets, with multi-site glycoproteins containing multiple compositions at each site. As a result, if the entire sample were to be injected at once then the mass spectrum would contain far too many peaks to be interpretable, and there would be too many precursors to analyse by MS2 at the same time. As such, prior to analysis by mass spectrometry, a separation stage is necessary. Liquid chromatography (LC) is commonly used as a separation stage in combination with electrospray ionization as it enables the in-line separation of glycopeptides followed immediately by mass spectrometry analysis. The LC method herein used for site-specific analysis is termed reverse phase LC, whereby a hydrophobic stationary phase, in this case an octadecyl carbon chain (C18) fused to silica, interacts with the hydrophobic

regions of the glycopeptides. More hydrophilic molecules will form weaker interactions with the stationary phase inside the column and will elute first. To maximise separation, a gradient of two solvents is used, beginning with a high concentration of polar solvents such as water, and then increasing the concentration of an organic (non-polar) solvent, such as acetonitrile, during the experiment [158].

As glycopeptides elute from the column they are then ionized via a similar mechanism as outlined above and the ions enter the mass spectrometer. The Orbitrap Fusion mass spectrometer uses similar, but distinct in places, methods to determine the *m/z* and fragment ions of interest. A schematic of this instrument is shown in **Appendix C.** For both MS1 and MS2 detection in this thesis an Orbitrap mass analyser was used, instead of time of flight as is present in the Synapt G2Si. An orbitrap consists of an outer barrel-like electrode and a coaxial inner spindle-like electrode which traps ions in an orbital motion around the central spindle [159]. As ions enter the orbitrap they begin to orbit the central electrode and induce a current in the outer electrode. The angular frequency (how quickly the ions are orbiting) of the ions is related to the *m/z* of the ion, and using a Fourier transform function it is possible to convert the frequency into a *m/z* readout. An orbitrap mass analyser enables high resolution detection, which is important for the accurate assignment of *m/z* for glycopeptide ions. Following initial MS1 detection, a quadrupole selects precursors of interest and fragments them using high-energy collision induced dissociation (HCD) [160]. The *m/z* of the fragment ions are then determined using the orbitrap. During MS2 fragmentation both the peptide backbone and the attached glycan are broken apart. In HCD fragmentation three kinds of peptide fragmentation can occur, resulting in the generation of a, b, and y ions (**Figure 1-10**). Using HCD fragmentation the N-linked glycan is itself fragmented from the peptide backbone and as such the peptide fragments can be used to sequence the glycopeptide of interest. This can then be searched against a protein sequence of interest to determine which PNGS this glycopeptide covers.

Figure 1-11 details this process for one glycopeptide. This peptide has one N-linked glycosylation site and will have different N-linked glycans attached, forming different *m/z* peaks. From the LC chromatogram (**Figure 1-11A**) two glycopeptides with different glycan compositions can be seen to elute at different times. As they elute from the column their *m/z* is detected. The first peak to elute has an *m/z* of 1020.4 (**Figure 1-11B**). The instrument will then isolate only ions with *m/z*=1020.4 and fragment them. The detected fragments are then used to confirm the amino acid sequence and to assign a particular glycan to that peptide (**Figure 1-11C**). With the fragmentation data and the precursor *m/z* assigned, it is possible to assign this glycopeptide with a $Man_8GlcNac_2$ glycan. Another glycopeptide with a larger *m/z* (**Figure 1-11D**) eluted from the column after this. The same analysis is applied to this peptide (**Figure 1-11D** and **4E**) and the glycan attached here is shown to possess

an additional mannose residue. All of the different glycan compositions for each peptide with the same amino acid composition then have their relative amount determined using the extracted ion chromatogram (XIC), which displays at what time during an LC run a particular ion was detected, and the intensity of that signal. This enables the relative amounts of different glycoforms to be determine at that site. For this example site, only a few oligomannose-type glycans were detected (**Figure 1-11F**). As certain N-linked glycan sites can be heterogeneous, glycans are often categorized according to their processing state and the presence and absence of fucose (**Figure 1-11F**). These methods form the basis of the glycan analysis performed in this thesis and the combination of released glycan analysis and site-specific glycan analysis enable the in-depth determination of the glycosylation state of viral spike proteins and can be applied to any spike protein of interest, with minimal adaptation.

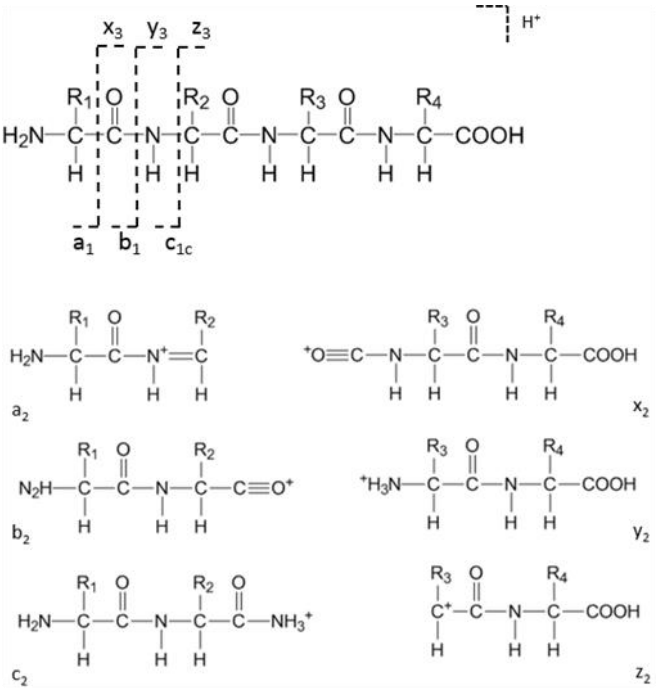


Figure 1-10: Peptide fragmentation patterns resulting from HCD fragmentation used in site specific glycan analysis workflows by LC-MS.
Schematic of potential peptide fragmentation of a peptide consisting of four amino acids with side chains R1, R2, R3 and R4 [161]

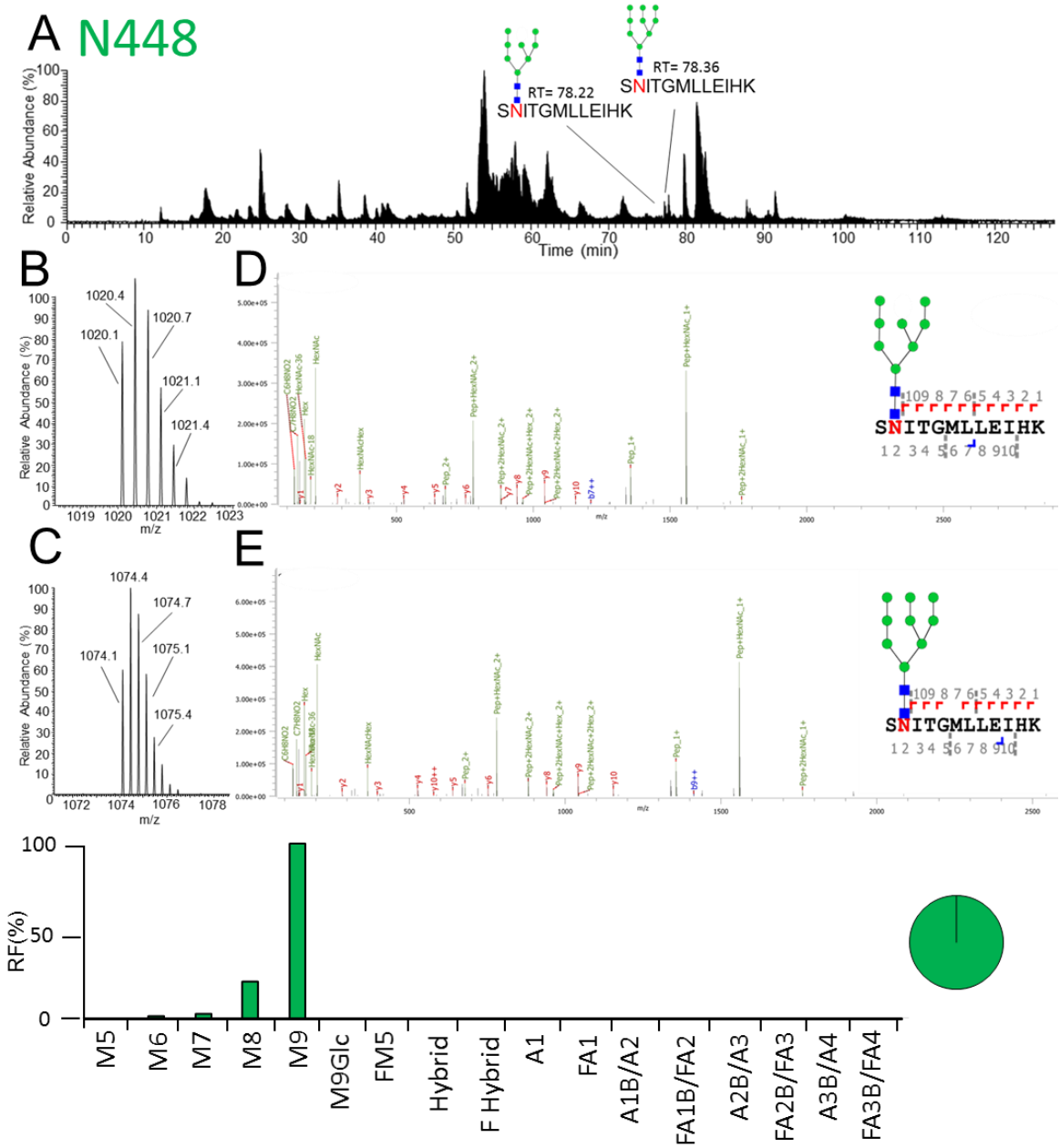


Figure 1-11: Workflow for analysing intact glycopeptide site-specific data.
(A) The LC chromatogram of Env digested with trypsin. Two peptides containing the N448 glycan site are annotated. (B) Extracted MS1 of the glycopeptide corresponding to the peak eluted at 78.22 minutes. (C) Extracted MS1 of the glycopeptide corresponding to the peak eluted at 78.36. (D) MSMS spectrum for the glycopeptide depicted in (B). Fragments containing a glycan component are labelled green, with peptide b and y ions depicted in blue and red respectively. The fragmentation spectrum combined with the *m/z* confirms the presence of a $Man_8GlcNac_2$ glycan. (E) MSMS spectrum for the glycopeptide depicted in C labelled in the same way as (D). The mass shift and fragmentation indicate the presence of a $Man_9GlcNAc_2$. The bottom bar graphs depict the relative amounts of each category of glycan, as determined by the XIC area for each glycan variant. Oligomannose-type glycans are depicted in green and complex-type glycans in pink, with a pie chart displaying the total amounts of each type. This format is used for every site to determine the site-specific glycan content at each site.

1.7 Site-specific characterisation of Env immunogens

The application of the glycan analysis workflow previously outlined to the recombinant Env mimic BG505 SOSIP.664 enabled Behrens *et al.* to establish several parameters that define native-like Env glycosylation that can be applied across different Env immunogens. Site-specific analysis upon trimeric BG505 SOSIP.664 revealed that a large number of sites across Env present only oligomannose-type glycans (**Figure 1-12A**). Sites that present complex-type glycans and a mixture of complex-type and oligomannose-type glycans are also present on gp120 and gp41. To better visualize the location of the oligomannose glycans Behrens *et al.* modelled the predominant glycan detected at each site onto the cryo-EM structure of BG505 SOSIP.664 (**Figure 1-12B**) [100].

Oligomannose-type glycans were shown to cluster at two regions of the trimer, on the outer domain of gp120, centred around N332 and N339, and towards the apex of the trimer at N160 and N156. The complex type glycans are located towards the base of the trimer and surrounding the mannose patch located at the apex of the trimer. A subsequent comparative analysis of trimeric BG505 SOSIP.664 with monomeric gp120 revealed that the mannose patch located at the apex of gp120 at N156 and N160 was only present when BG505 SOSIP.664 gp120 was expressed as a trimer (**Figure 1-12C**) [79]. This led to the definition of two different mannose patches on Env. The first is termed the intrinsic mannose patch (IMP), located around the N332 glycan, as it is present on gp120 monomer and trimer. The other mannose patch focused around N160 and N156 was termed the trimer associated mannose patch (TAMP) (**Figure 1-11D**). Subsequent analyses from other groups have confirmed these phenomena and these regions of oligomannosylation are key indicators of natively glycosylated Env as they form key bnAb epitopes [162,163].

These studies revealed that glycan processing on viral spike proteins can be distinct from those found on typical glycoproteins located on the cell surface. Behrens *et al.* further explained that the reason for the high abundance of oligomannose-type glycans on Env are present due to the density of the glycan shield of Env [118]. The high number of PNGS on BG505 mean that there are steric constraints placed upon the ER and Golgi mannosidase enzymes, and they cannot act upon the viral glycans, leaving abnormal oligomannose-type glycans on the mature protein. To visualize this effect, the crystal structure of the catalytic domain of ER-α mannosidase I was docked onto key glycan sites within the IMP [118]. Whilst this only served as a cartoon representation and was not calculated from molecular dynamics simulations, it is clear that the active site of the mannosidase enzyme cannot access the N332 glycan as other proximal glycans clash with the mannosidase structure (**Figure 1-13**).

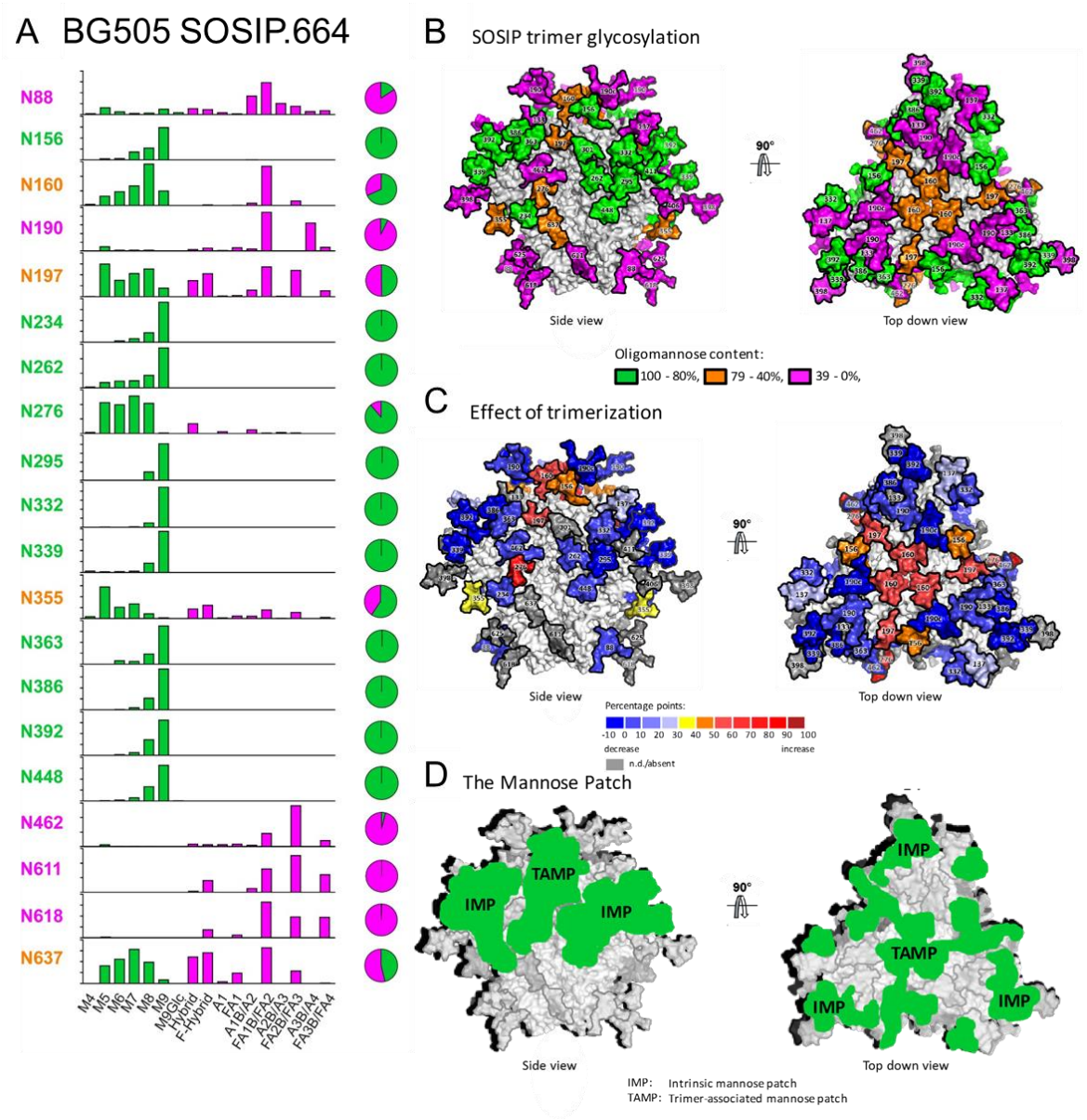


Figure 1-12. Summary of the site-specific analysis of BG505 SOSIP.664, defining key features of Env glycosylation.
(A) Site-specific glycan analysis of BG505 SOSIP.664 reproduced from Behrens *et al.* 2016 displaying categories utilised for subsequent site-specific analysis [100]. Oligomannose-type glycans are shown in green and complex in magenta. (B) The predominant glycan detected at each site was modelled onto BG505 SOSIP.664 Env cryo-EM density (PDB ID: 5ACO) and coloured by the amount of oligomannose detected at each site from (A). (C) Using the model in B the changes in oligomannose content at each site between trimeric gp120 and monomeric gp120 were mapped enabling the visualization of glycans that are protected from processing when Env is expressed as a trimer [79]. (D) Mapping the two regions of under processed oligomannose-type glycans that are key features of native-like Env trimers. Panels B, C and D are reproduced from Behrens *et al.* 2017 [118].

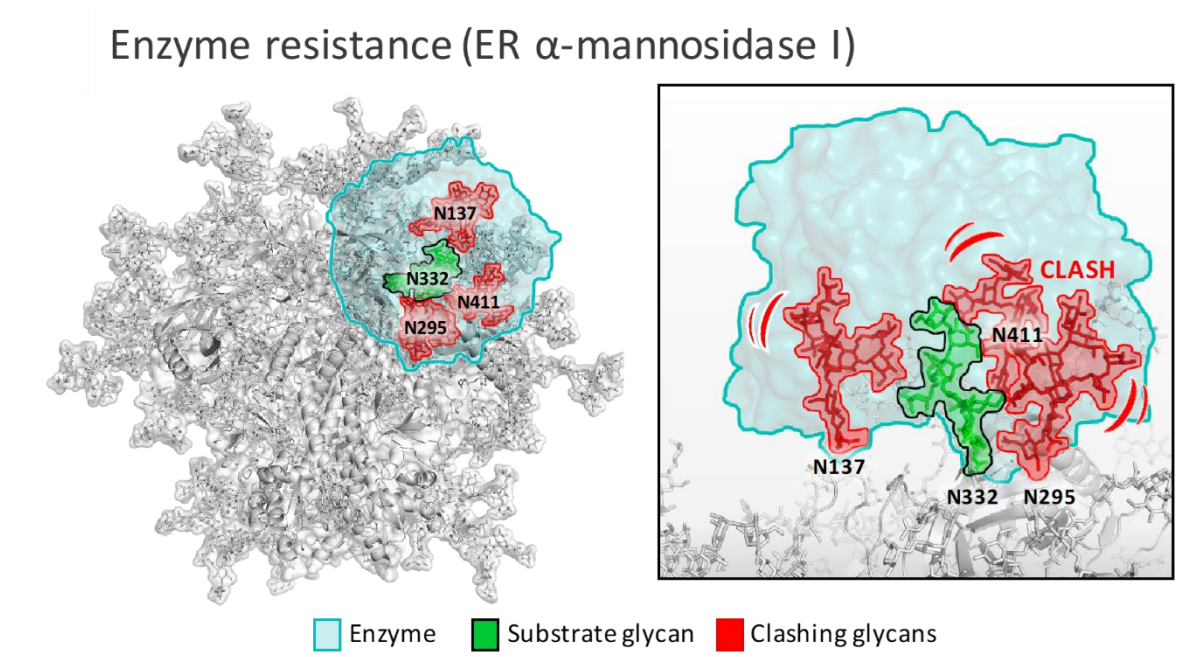


Figure 1-13: Diagrammatic representation of the steric clashing of neighbouring N-linked glycans preventing glycan processing at N332.
(A) ER α1,2-mannosidase I (PDB ID 5KIJ) [164] (cyan) was modelled to bind the $Man_9GlcNAc_2$ N332 glycan (green) on a fully glycosylated model of a previously described model of BG505 SOSIP.664 [100]. The surrounding glycans that sterically clash with enzyme recognition of N332 are highlighted in red. Figure reproduced from Behrens *et al.* 2018 [118].

This principle was further demonstrated with the analysis of glycan depleted BG505 (**Figure 1-14A)**. For these trimers, glycan sites were removed to promote antibody binding, and this had a knock-on effect on the glycan processing of adjacent sites [165]. One example of this is N234, which forms part of the IMP. When the adjacent N-linked glycan site N276 was removed, $Man_9GlcNAc_2$ was depleted from the N234 site (**Figure 1-14B**), likely due to the removal of clashes between the N276 site and the ER α1,2-mannosidase I enzyme, enabling the N234 glycan to be processed. This shows how the glycan shield of Env forms an interconnected network that disrupts and controls N-linked glycan processing.

Understanding the glycosylation of viral glycoproteins is a key stage in the vaccine design process, as it defines a large part of the antigenic surface of the target of interest. This is especially true for HIV-1 as the glycan shield is the target for a number of potent broadly neutralizing antibodies. These are thought to be key to eliciting a protective response against HIV-1 infection, and as such mapping the epitopes of carbohydrate binding antibodies is a key application of the methodologies outlined above. Furthermore, for emergent viruses such as SARS-CoV-2 it is important to gain as much knowledge about the structure of the protein to maximise vaccine success, and also to establish whether the glycosylation patterns of HIV-1 Env are generic, or only applicable to Env.

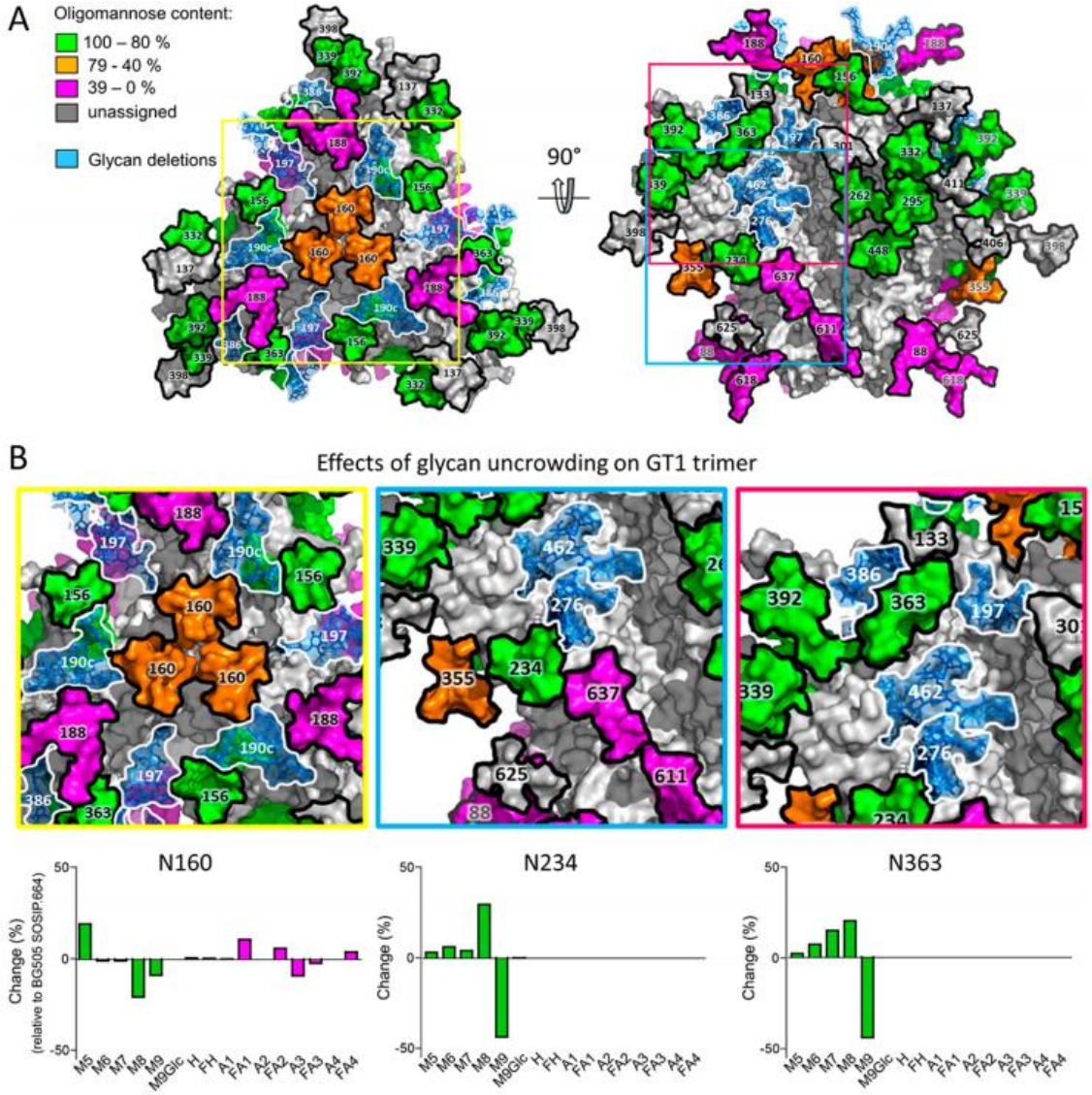


Figure 1-14: Processing of glycans on the glycan depleted BG505 trimer.
(A) Model is based on the one presented in **Figure 1-12**, but with the glycans deleted from the trimer shown in light blue with a white border. The remaining glycans are colour-coded according to their processing state (see key on the figure). (B) Magnification of three sites (N160, N234, and N363) to highlight the effects of the reduction in local glycan density on the trimer. Underneath these images are difference plots of the % change in glycan abundances for the glycan depleted trimer compared with SOSIP.664. Reproduced from Behrens *et al.* 2018 [165].

1.8 Vaccination strategies involving glycoproteins to protect against HIV-1 and SARS-CoV-2.

1.8.1 How a vaccine initiates an antibody response

A vaccine is described as a biological product that can safely induce an immune response to a particular disease prior to infection. Vaccines, in one form or another, have been around for hundreds of years and have increased in sophistication as time has passed [166]. The first example of immunization against smallpox was variolation, in which live smallpox virus was placed into the skin. This resulted in a more localized infection compared to inhalation, but still provided protection against subsequent infection. The first vaccine for smallpox involved administering a similar, but much less lethal, cowpox virus that provided protection from subsequent exposures to smallpox [167]. Since then, the incidence of a number of diseases has been cut drastically, and smallpox has been eradicated [168]. Vaccines remain the most successful tool to control epidemics in the long term. The biological product that confers immunity in a vaccine comes in many forms: attenuated or inactivated, whereby the infectivity/ severity of infection of a particular pathogen is compromised prior to injection, or through the administration of part of the pathogen in a protein subunit, viral vector, or nucleotide-based approach, whereby the immune response to the portion of the pathogen is sufficient to protect against infection of the real virus. HIV-1 vaccines have tended towards subunit approaches, as either live-attenuated or whole-inactivated viruses could not be used in HIV vaccine development due to the danger associated with integration of the viral DNA in the host chromosome [169]. For SARS-CoV-2 vaccine design, a broad range of approaches have been applied, including inactivated virus and subunit vaccines delivered using a nucleotide-based approach, which is detailed in **Chapter 1.8.4**.

A subunit vaccine consists typically of the immunogen, an adjuvant, and other components, such as emulsifiers, to aid with solubility and stability. An adjuvant is an additive that enhances the immune response towards a vaccine, typically through activating the innate immune system. Alum is a common adjuvant, however it's exact mechanism of action still remains unknown [170]. A typical immune response to a subunit vaccine administered intramuscularly is shown in **Figure 1-15**. Once administered, the protein is taken up by antigen presenting cells (APCs), especially dendritic cells, which are themselves activated through pattern recognition receptors (PRRs) recognizing the danger signals provided by the adjuvant [166]. Following antigen uptake and activation, the APC is trafficked to the draining lymph node. The engulfed protein is digested, and peptides are then presented on the surface of the APC by major histocompatibility complexes (MHCs).

These presented peptides activate T cells via the interaction of the MHC-peptide with the T cell receptor (TCR) [171]. In addition to this process, the vaccine immunogen has also trafficked to the lymphatic system, and these are recognized by B cell receptors. The activated T cells then promote B cell development in the lymph node, termed CD4+ T help.

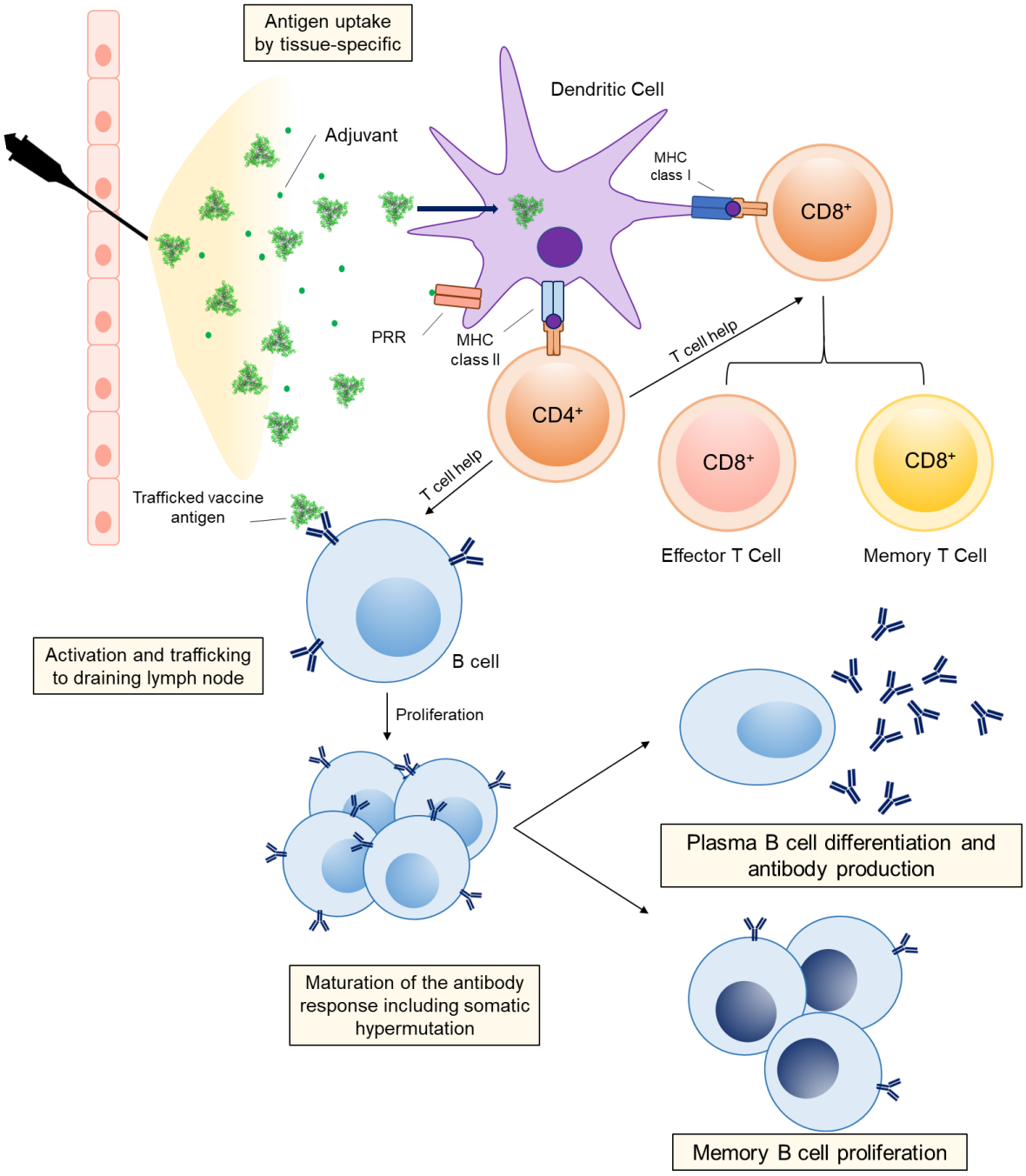


Figure 1-15: Summary of the immune response resulting from immunization with a protein subunit vaccine.

This activation results in the maturation of the antibody response, through processes such as somatic hypermutation, to improve antibody affinity and induce isotype switching [172]. These short-lived plasma cells actively secrete antibodies specific for the immunogen administered, and the serum antibody levels also rise. In addition to this, memory B cells are produced, which provide immune memory, and are rapidly activated if a similar antigen

is encountered again. Long lived plasma B cells can continue to produce antibodies for decades, and reside in the bone marrow [173]. In addition to CD4+ T cell help, vaccines can also induce CD8+ T cells, which perform a range of functions, including the destruction of infected cells presenting pathogen derived peptides via MHC molecules [174]. The combination of B and T cell mediated immune memory via memory B and T cells should then be capable of inducing a strong adaptive immune response in a much faster time upon encountering an infectious agent, without requiring the process of affinity maturation.

1.8.2 **The goal of vaccine design for HIV-1 Env, broadly neutralizing antibodies.**

The HIV-1 Envelope glycoprotein is one of the most densely glycosylated proteins in nature, with half of the total protein mass consisting of carbohydrates [175]. As HIV-1 hijacks the hosts N-linked glycan pathway Env recruits around 100 N-linked glycans across the Env trimer. These carbohydrates are thought to shield the underlying protein surface preventing the immune system from recognising the highly antigenic viral protein surface [176]. Paradoxically, however, potent and broad neutralising antibodies have been isolated from infected patient sera that have been shown to interact with this coat of host-derived sugars [85]. Although protein-carbohydrate interactions are much weaker than protein-protein interactions, an increasing number of glycan binding antibodies are being isolated and the remarkable breadth of their response and the ability to protect against HIV-1 infection via passive transfer, form the focus of HIV-1 vaccine design [85,86,88].

The potency and breadth of these antibodies can be rationalised when considering the conservation of the N-linked glycosylation sites across many strains of Env (**Figure 1-4**). The first broadly neutralising antibody identified, 2G12, possesses an epitope that consists entirely of glycans, recognizing under processed oligomannose-type glycans focused around the intrinsic mannose patch [177]. This discovery stimulated both vaccine design efforts and the search for more such bnAbs. Screening infected patient sera has resulted in the characterisation of a plethora of bnAbs [81,178–180]. Encouragingly these antibodies bind across the regions of the trimer and to a range of different types of N-glycans. Although 2G12 remains the only bnAb with an epitope consisting solely of glycans, it is not the only one to recognise oligomannose-type glycans [181]. Such an extensive number bind to a region of dense oligomannose-type glycans centred around N332 that this site has been termed the “supersite of vulnerability” [182], it is therefore essential to ensure that immunogens

present oligomannose-type glycans at this site. One such antibody is PGT135, which binds to the oligomannose-type glycans found at N332, N386 and N392 (**Figure 1-16A**) [179,181].

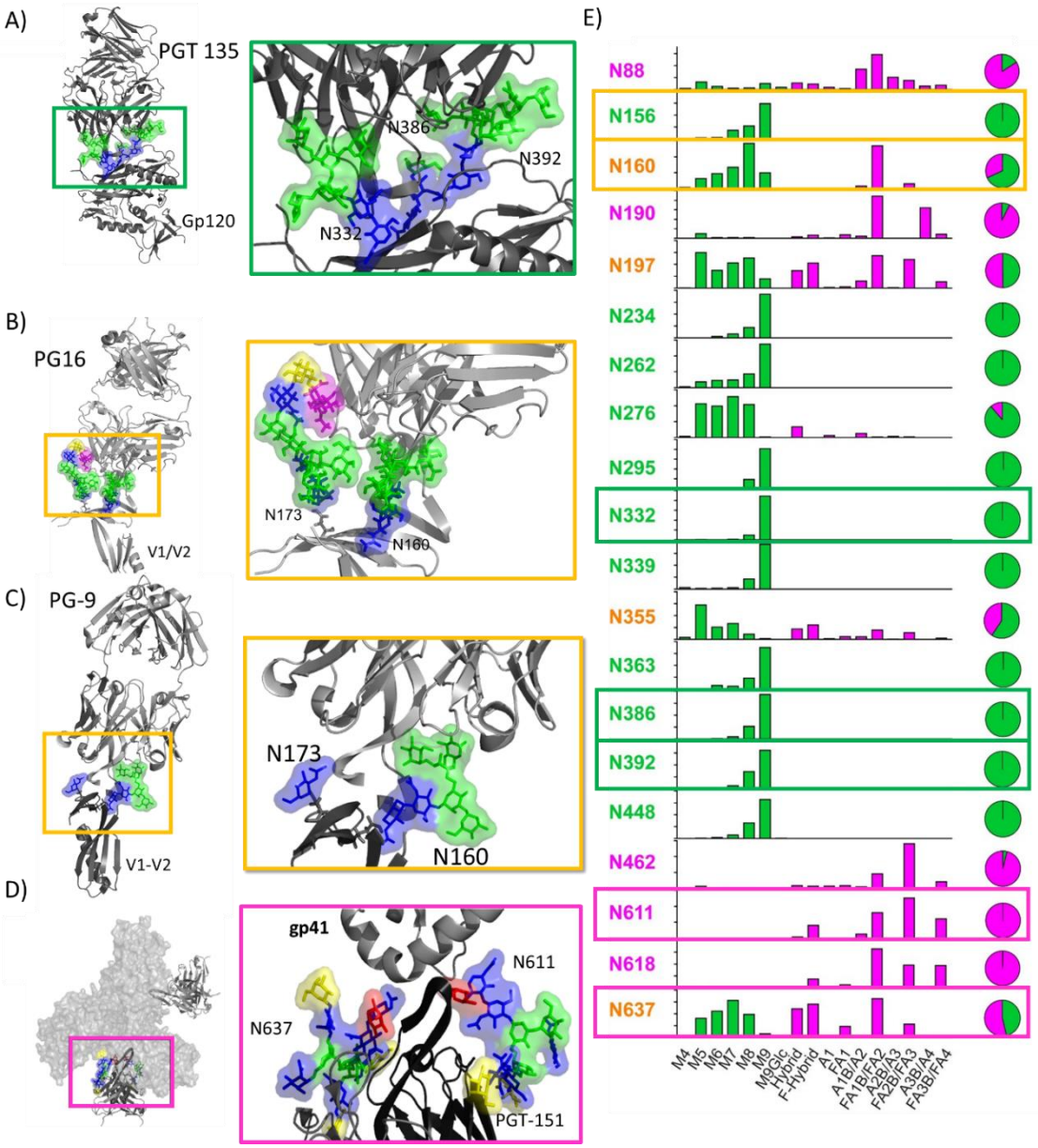


Figure 1-16: Glycan epitopes of broadly neutralizing antibodies.
(A) Crystal structure of PGT-135 with glycans that interact with the PGT135 highlighted and the residues to which the glycans are attached labelled (PDB ID: 4JM2) [179]. N-glycan monosaccharides are coloured blue (GlcNAc), green (mannose), fucose (red), galactose (yellow) and sialic acid (purple). (B & C) Interaction of PG16 and PG9 with Env N-glycans, labelled and represented as outlined in panel A (PDB ID: 4DQO/3U4E) [183,184] (D) A cryo-EM representation of PGT151 in complex with JRCSF trimer. The expanded region highlights the complex glycans on N611 and N637 (PDB ID:5FUU) [83] (E) Site-specific analysis of SOSIP adapted from Behrens *et al.* (2016) [100] with the residues highlighted in yellow from PG16, green for PGT135 and orange for PGT151. Oligomannose shown in green, complex/hybrid in magenta [100].

Epitopes for broadly neutralising antibodies are not limited to this region. Apex-binding broadly neutralising antibodies have been characterised that also recognise under processed oligomannose-type glycans and are extremely potent. Two of such antibodies are shown in **Figure 1-16B** and **C**; PG9 and PG16. Both recognise oligomannose-type glycans at N160 with PG16 also displaying a preference for hybrid-type glycans [184]. The ability of the gp120-41 interface binding bnAb PGT151 to recognise complex-type glycan structures exemplifies the ability of bnAbs to recognise all different kinds of N-linked glycans [107]. It is therefore of paramount importance for candidate immunogens to have their entire glycan surface characterised to ensure that the oligomannose, complex and hybrid-type glycans are presented in the appropriate manner on a site-specific level.

Despite the characterisation of bnAbs that recognise the entire glycan surface and the development of Env trimers that closely mimic full length Env, in the form of SOSIP.664, a successful strategy to elicit broadly neutralising antibodies has not currently been achieved. This is a result of the large number of somatic hypermutations required for HIV-1 bnAbs to accommodate glycans as part of their epitopes. Several studies have also revealed the rarity of the precursor B cells that could potentially evolve to become bnAb producing B cells in naïve individuals [185]. It was previously thought that the application of a native-like HIV-1 trimer would be sufficient to produce a strong response. As such it is important to understand how a vaccine induces an immune response to develop strategies to encourage the immune system to produce bnAbs.

1.8.3 **Vaccination strategies against HIV-1 to induce bnAbs**

Overcoming the metastability of Env to generate stable soluble immunogens, such as BG505 SOSIP.664 catalysed a new wave of immunogen design. With the ability to generate native-like Env-based immunogens, combined with the discovery of many broadly neutralizing antibodies that could protect against infection if elicited prior to infection, global vaccine design efforts shifted towards an immunogen, or series of immunogens, that can elicit broadly neutralizing antibodies. Whilst studies in animal models, including macaques, have shown that BG505 SOSIP.664 alone is insufficient to elicit a bnAb based response [186], these are only animal models and as such clinical trials are ongoing to establish the human immune system's response to native-like trimers in the form of BG505 SOSIP.664 [187]. Studies in animal models demonstrated that when BG505 SOSIP.664 is administered, neutralizing antibodies are elicited which are described as autologous, which means that they are only capable of neutralizing the BG505 strain, which is not sufficient for HIV-1 vaccines due to the multitude of quasi-species circulating in an infected individual [66,186].

Novel approaches to investigate how bnAbs arise during natural infection involve deep sequencing the antibody response in infected sera and then studying the evolution of a particular bnAb from the baseline human antibody repertoire [188]. The antibodies produced by these evolutionary precursors to bnAbs are referred to as inferred germline (iGL) antibodies, and the approach of reverse vaccinology refers to developing immunogens capable of binding to the iGL antibodies, with the goal of beginning the evolution of the iGL B cells to produce bnAbs [188]. Native-like immunogens are incapable of binding to iGL antibodies and as such bespoke antigens are required in this approach.

An example of germline targeting is to kickstart the development of bnAbs targeting the CD4 binding site (CD4bs). This region of Env is a promising target, as studies have shown that bnAbs that bind to this site are extremely broad and potent [189]**.** One approach involves extracting a portion of gp120 containing the CD4 binding site and expressing this region alone. These constructs are referred to as engineered outer domain (eOD) immunogens [190]. The lack of other regions of Env on these immunogens reduces the number of undesirable epitopes that may be immunodominant but not able to elicit bnAbs. These immunogens are designed to prime the precursors for the VRC01 bnAb, an extremely potent and broad antibody that recognizes the CD4 binding site (CD4bs) [191]. This bnAb does not interact directly with a glycan, instead VRC01 pushes the conserved N276 glycan aside to contact the CD4bs. It is still important to characterise the glycans present on eOD trimers to ensure that the protein surface remains covered to focus the immune response towards the CD4bs. To present a larger number of eOD particles to the immune system eOD monomers have been assembled into nanoparticles containing 60 subunits, termed the eOD 60mer. These constructs are currently being produced according to good manufacturing practises, and the resultant material is being used in a phase 1 clinical trial in humans [192]. Another platform for CD4bs germline targeting is the glycan depleted trimer, termed BG505 SOSIP.v4.1-GT1 which has been designed to trigger B cells corresponding to germline PGT121 and germline VRC01. This trimer lacks 15 glycans across the viral spike and initiates antibody responses in knock-in mice expressing the predicted germlines for these bnAbs [193]. The glycan analysis of this immunogen is described in **Figure 1-14**, and revealed that the glycan shield was largely unaffected by the glycan deletions with changes in their mannose trimming localised to regions proximal to a depleted glycan [165].

To maximise the success of any HIV-1 vaccine candidate it is important to characterize the immune response, minimize any unwanted narrow autologous responses towards the trimer, and to establish the parameters that gave rise to autologous responses. This is where glycan analysis is critical, as the majority of the surface of Env is glycosylated, and so comparing which glycans are present on candidate immunogens, compared to Env

displayed on an infectious virion, is key to ensure neutralizing, and non-autologous, epitopes are conserved.

1.8.4 Vaccine design in the era of COVID-19, rapid development, testing and deployment

Unlike HIV-1, the majority of infections are cleared by the host prior to death. The death rate of COVID-19 in under 50s is low, however, due to the high infectivity of the virus the low death rate is magnified into millions of deaths. A vaccine against COVID-19 must reproduce, or improve upon, the immune response to infection. In response to COVID-19 infection, the immune system is capable of inducing a strong, long lived, neutralizing antibody response [57]. Patient sera from infected individuals were rapidly obtained by multiple labs and the antibody response was mapped onto the S protein surface. In patients from a cohort in Amsterdam, the strongest neutralizers bound to the RBD [194] (**Figure 1-17A**). Several studies supported this observation from across the globe, and it is clear that the RBD of SARS-CoV-2 is an immunodominant epitope [195–205]. Additionally, these studies showed that the antibody response elicited against the RBD was diverse, and that a large number of different heavy chain genes can be used by the humoral immune system to target the RBD (**Figure 1-17B**). This is in contrast to HIV-1 Env bnAbs where only a limited subset of heavy chain genes can evolve to generate bnAbs [206].

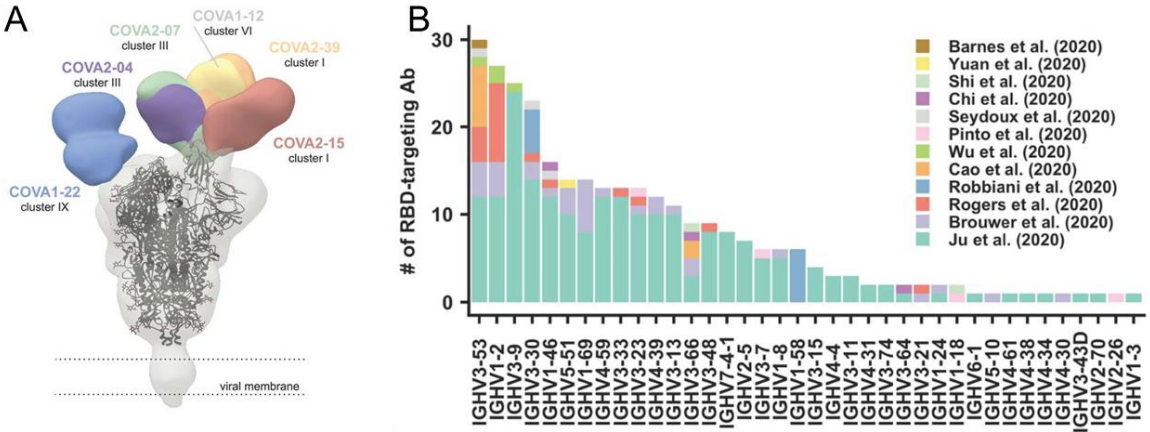


Figure 1-17: Studies from infected patient sera published during initial COVID-19 outbreak in 2020 highlighting the propensity to target the RBD of the spike protein. (A) Figure panel produced from Brouwer *et al.* highlighting example anti-SARS Abs that recognize the RBD [194]. Composite figure demonstrating binding of NTD-mAb COVA1-22 (blue) and RBD mAbs COVA2-07 (green), COVA2-39 (orange), COVA1-12 (yellow), COVA2-15 (salmon), and COVA2-04 (purple) to SARS-CoV-2 spike (grey). The spike model (PDB 6VYB) is fit into the density. (B) Panel reproduced from Yuan *et al.* 2020 highlighting the large number of Abs elicited against the RBD of SARS-CoV-2 during infection [197]. The distribution of Immunoglobulin heavy chain variable (IGHV) gene usage is shown for a total of 294 RBD-targeting antibodies [195–205].

The robust adaptive immune response to SARS-CoV-2 infection is promising for vaccine design, however due to the rapid onset and spread of the COVID-19 pandemic novel vaccine strategies were required to test and manufacture large scale quantities in short amounts of time. As with HIV-1, vaccine design efforts have focused around the presentation of the spike protein, but the method of delivery involves new and previously untested formats. Typically, glycoproteins for use as a therapeutic, or as a vaccine, are made recombinantly using Chinese Hamster Ovary (CHO) cells, where the gene of the target of interest has been introduced to create clones [207]. Rounds of clonal selection are performed to select for parameters, including the amount of protein generated and many other key characteristics. In addition, as the protein is being used for human trials there are a multitude of further quality control checkpoints required to ensure that the final protein product is safe for use in humans. These steps are costly, and it is very challenging to obtain the yield and purity needed for Good Manufacturing Practise (GMP) certification. This was successfully performed for BG505 SOSIP.664, BG505 SOSIP.664-GT1 and eOD GT8 [109,208,209], as well as several trimers developed by the European AIDS vaccine initiative, however this took a number of years to achieve. For SARS-CoV-2, vaccine design approaches were investigated that could bypass the need for protein production *in vitro* and therefore avoid the costly and lengthy process of generating GMP-grade material.

Instead of recombinant protein immunogens, examples of delivery mechanisms for SARS-CoV-2 vaccines include mRNA (used in the Moderna and BioNtech/Pfizer). adenovirus (Astra Zeneca) and inactivated virus (Sinovac) [210–217]. In the mRNA and adenovirus-based approaches, genetic material is administered that results in the production of immunogen by the recipients own cells, in both cases the SARS-CoV-2 S protein. As such, even though the vaccines do not contain glycans themselves, the vaccine is inducing the production of glycosylated SARS-CoV-2 spike protein, and therefore glycan analysis of SARS-CoV-2 spike protein is important for understanding the antigenic surface of the SARS-CoV-2 spike protein.

1.9 Applying glycan analytics to inform the immunogen design process.

Employing and expanding upon the strategies outlined in **Chapter 1.6,** this thesis aims to build upon prior knowledge of the presentation and processing of N-linked glycans on candidate immunogens and their viral counterparts. Prior to the execution of this work, no study had been performed comparing a sequence matched soluble native-like trimer of HIV-1 Env (BG505 SOSIP.664) to the corresponding virus, BG505 T332N. Factors including the choice of cell line for expression and the presence or absence of the transmembrane region of Env all influence the resultant glycosylation of the mature

glycoprotein [103]. This work directly compared the glycosylation of Env derived from the BG505.T332N-LAI virus, produced in a human lymphoid cell line, with the corresponding BG505 SOSIP.664 soluble trimers expressed in CHO or 293F cells. The differences in producer cell glycosylation are important as transient transfection using HEK293F cells, used for Env production in the lab, are not typically used in the GMP production of candidate vaccines. The batch-to-batch variability resulting from transient transfection, whereby cells are cultured and then transfected with the gene of interest, is not commonly used due to poor consistency. It is much more common to introduce the gene of interest, in this case BG505 SOSIP.664, into the CHO cells of interest, and then clonally select the clones that have the best yield [109]. By comparing material produced recombinantly in a research laboratory and GMP-grade, with viral derived Env it will be possible to identify any differences in the glycan shield and therefore differences between the antigenic presentation of the virus and the vaccine candidate. This work is presented in **Chapter 3**.

This comparative analysis revealed that whilst the majority of the glycan shield was conserved, regarding the presentation of oligomannose-type glycans at the IMP and TAMP, there were localized differences between recombinant and viral-derived Env. One observation was that the glycan occupancy of particular N-linked glycan sites on BG505 SOSIP.664 was less than that of the viral counterpart. Concurrent studies by other laboratories demonstrated that these artificial holes in the glycan shield form an immunodominant epitope, whereby an antibody response is raised to the glycan hole, but the elicited antibody cannot neutralize the autologous virus, unless a particular glycan site is knocked out [218]. This important difference between recombinant protein and virus requires addressing to improve the immune response to vaccine candidates. To this end, **Chapter 4** describes a joint project taken place during a collaborative visit to the University of Southampton in 2018 with Dr Ronald Derking, based in the laboratory of Prof. Rogier Sanders (University of Amsterdam), who designed the original BG505 SOSIP.664. Glycan methods used for previous studies were adapted to include site-specific glycan occupancy to demonstrate the presence of artificial glycan holes on BG505 SOSIP.664. It was observed that, for BG505 SOSIP.664, the sites that displayed reduced site-specific glycan occupancy consisted of NxS codons. Previous literature reported that NxT codons have higher affinity to the glycan attachment machinery, and as such NxT sequons replaced NxS [155]. Glycan analysis and antibody binding assays demonstrated that the artificial glycan holes had been filled, demonstrating a mechanism to produce SOSIP immunogen candidates with a more viral-like glycan shield.

The application of glycan analysis methodologies to study HIV-1 were employed to investigate SARS-CoV-2, as viral infections reached pandemic levels in early 2020.

Analysing the glycan shield of SARS-CoV-2 S protein is beneficial for the vaccine design process, but additionally it enables further investigation into whether the constraints on glycan processing are conserved on other viruses, or if they are unique to HIV-1 Env. Using stabilized recombinant S protein in collaboration with laboratory of Prof. Jason McLellan (University of Austin, Texas), who first reported the structure of the SARS-CoV-2 spike protein [69], this thesis reports the first characterization of the glycan shield of the SARS-CoV-2 spike protein in **Chapter 5**, and demonstrated similar glycan signatures as on HIV-1 Env with the presentation of oligomannose-type glycans. However, the abundance of oligomannose-type glycans on S protein was reduced. This reflected the denser glycosylation of HIV-1 Env compared to S protein and highlights one of the reasons why it is possible for most SARS-CoV-2 infections to be cleared by the immune system and not for HIV-1.

As with HIV-1 Env it is important to investigate the presentation of the glycan shield across multiple cell lines and compare the glycosylation of recombinant and viral SARS-CoV-2 S protein. In **Chapter 6** a comparative analysis of SARS-CoV-2 S protein is performed on recombinant S protein produced from multiple laboratories across the globe, including material produced from HEK293F cells and CHO cells, with data obtained from viral-derived S protein. Unlike HIV-1 Env, the glycosylation of SARS-CoV-2 S protein was consistent in both the presentation of oligomannose-type glycans and glycan occupancy, and as such the glycan shield of SARS-CoV-2 S protein is not hindering the efficacy of vaccines in the same way as for HIV-1 Env. Finally, the analytical workflows established to analyse the envelope glycoproteins of HIV-1 and SARS-CoV-2 was employed to investigate whether the glycans on the ACE2 receptor influence the binding of SARS-CoV-2. **Chapter 7** explores whether there is a role for glycosylation in SARS-CoV-2 pathobiology beyond shielding epitopes of interest. By combining site-specific glycan analysis with surface plasmon resonance (SPR) and glycan engineering, it was shown that the specific composition of the glycans on ACE2 impact binding to S protein, however these changes are subtle, and therefore glycan interactions are likely not critical for SARS-CoV-2 infection. These analyses demonstrate the differential importance of N-linked glycosylation in vaccine design, with critical differences in the glycan shield of HIV-1 Env between recombinant and viral Env resulting in the presentation of undesirable antibody epitopes, compared to the similarity of recombinant and viral derived S protein.

Chapter 2 Materials and Methods

2.1 Materials and methods Chapter 3

2.1.1 Production of BG505.T332N-LAI gp120 and gp41 (from the laboratory of Prof. Lifson, Frederick National Laboratory, US)

An infectious stock of HIV-1 BG505.T332N-LAI was derived by transfecting 293T cells with the infectious molecular clone (IMC). The IMC Env glycoproteins have gp120 and gp41-ectodomain subunits derived from the primary R5 Tier-2, clade A BG505 week 6 isolate, but with the N332-glycan knocked-in to mimic the BG505 SOSIP.664 soluble trimer sequence [76,193]. The transfection stock was used to infect the CD4+, CCR5+ A66-R5 T cell line to provide the large amount of virus required for glycan analysis. Cytopathic effects were first observed in the culture on day-18 post infection, and progeny virus was first detected by p24 antigen capture immunoassay on day-21. Because the IMC was cytopathic, ~100-500,000 uninfected A66-R5 cells were added per mL of the culture at seven different time points during its 54-day duration, to maintain a high level of viability and sustain virus production. On day-54, 20 L of culture supernatant were harvested for virion purification by sucrose density-gradient centrifugation. The resulting 1,000-fold concentrated stock (lot P4408, 20 mL volume) of HIV-1 BG505.T332N-LAI/A66-R5 virus served as the starting material for purifying gp120 and gp41 Env glycoproteins. The mass amounts of p24 Gag protein and gp120 Env glycoprotein in the purified virus stock were estimated by analysing fluorescent dye-stained SDS-PAGE gel bands using a sensitive, densitometric technique, compared to standard curves derived using highly purified viral protein preparations; well characterized reference virus preparations (SIVmac239/SupT1-R5, HIV-1 BAL/SupT1-R5 and HIV-1 NL4-3/SupT1) served as additional controls.

2.1.2 Cell culture production of BG505.T332N-LAI virus. (from the laboratory of Prof. Lifson, Frederick National Laboratory, US)

The construction of the BG505.T332N-LAI IMC used in this study has been reported previously [193]. Neutralization assays using this virus were performed at the Academic Medical Centre, Amsterdam as previously described [219]. An infectious virus stock was prepared by transfecting 293T cells (American Type Culture Collection Cat. #11268) with the IMC using the TransIT®-293 transfection reagent as described by the manufacturer (Muris Inc). Aliquots of cell free culture supernatant (infectious stock) were collected and stored at -80˚C. To produce the BG505.T332N-LAI virus, a 50-mL starter culture of A66-R5

cells containing 500,000 viable cells/mL was inoculated with 1 mL of the above infectious stock. The culture was maintained in RPMI 1640 containing L-glutamine and supplemented with 10% foetal bovine serum (Gemini GemCell™ Lot A90D69E, heat inactivated) and antibiotics (1000 U/mL penicillin/1000 µg/mL streptomycin). Depending upon the cell density, the cultures were passaged at 1:2 to 1:5, twice weekly during scale up. Progeny virus production was monitored using an in-house HIV-1 p24 antigen capture immunoassay (AIDS and Cancer Virus Program, Frederick National Laboratory for Cancer Research).

2.1.3 Purification of BG505.T332N-LAI virus. (from the laboratory of Prof. Lifson, Frederick National Laboratory, US)

Supernatant from the infected A66-R5 cell culture was sequentially filtered to remove cells and large extracellular vesicles using 5.0 µm (Millipore Opticap® XL 10) and 0.5 µm (Millipore Opticap® XL 5) capsule filters, respectively. The viruses present were purified by continuous flow sucrose density gradient centrifugation followed by sucrose removal, and the final virus concentration was determined, all via procedures described elsewhere [220]. Aliquots of the purified virus (lot P4408) were stored in a liquid N2 vapor phase freezer prior to use.

2.1.4 Reversed-phase HPLC purification of virion gp120 and HPLC plus SDS-PAGE purification of virion gp120 and gp41. (from the laboratory of Prof. Lifson, Frederick National Laboratory, US)

The virus preparation (lot P4408) described above was disrupted in 8M Guanidine-HCl (Pierce, Rockford, IL) and the proteins fractionated by HPLC under non-reducing conditions to isolate the gp120 and gp41 components. HPLC was performed at a flow rate of 300 µL/min on a 2.1 x 100 mm Poros® R2/H narrow bore column (Applied Biosystems, Bedford, MA, USA), using aqueous acetonitrile/trifluoroacetic acid solvents and a Shimadzu HPLC system equipped with LC-10AD pumps, SCL-10A system controller, CTO-10AC oven, FRC-10A fraction collector and SPD-M10AV diode array detector. The gradient of buffer B (0.1% trifluoracetic acid in acetonitrile) was: 20%-36%, 5 min; 36%-43%, 14 min; 43%-50%, 11 min; 50%-80%, 5 min; and 80%, 5 min. A temperature of 55°C was maintained during HPLC separation. Protein peaks were detected by UV absorption at 206 and 280 nm. Fractions corresponding to gp120 and gp41 were subjected to SDS-PAGE. The gel bands corresponding to gp120 and gp41 were excised and stored at -20°C for use in UPLC and mass spectrometry analyses of released glycans.

A second aliquot of virus was processed by HPLC in an identical manner, resulting in the purification of ~70 µg of gp120 subunits, as quantified using the fluorescent dye staining

technique. Fractions containing the gp120 protein were collected, pooled, and lyophilized. This purified gp120 stock was used for subsequent site-specific glycan analysis.

2.1.5 **Production of BG505 SOSIP.664 trimers.**

Human Embryonic Kidney 293F cells (HEK293F) in a serum-free suspension culture, were grown in FreeStyle 293 Expression Medium (Invitrogen). A stock of these cells was made at the concentration of 1x105/ml. HEK293F cells double every 24 hours (range of 20-25 hours), are ready for transfection with a cell density of $1x10^6$/ml and should never exceed $3x10^6$/ml for subculturing. A small aliquot of the culture was taken and dyed with trypan blue to enable manual cell counting with a haemocytometer.

For transfection, the volume of culture used was 200ml. 1mg/ml of MaxiPrep DNA (200mg) in pre-warmed 1 in 40 dilution in culture (5ml) of OptiPro SFM medium (Invitrogen) was combined with 1µl/ml of culture of transfection reagent (200µl), 293fectin (Invitrogen), in 5ml OptiPro. After incubating for less than 10 minutes, the transfection mixture was added to $1x10^6$/ml HEK293F cells in 500ml Erlenmeyer flask with a vent cap (Corning). The cells were incubated at 37°C with 5% CO2, shaking at 137rpm as recommended by the manufacturer.

Correctly folded, cleaved BG505 SOSIP.664 trimers were purified from the supernatants using a Ni2+-NTA affinity column (GE Healthcare UK) followed by bnAb PGT151 affinity chromatography. The production and purification of the same BG505 SOSIP.664 trimers from a stable CHO cell line under cGMP conditions has been described elsewhere [109]

After harvesting, the cells were spun down at 4000 rpm for 10 minutes and the supernatant applied to a 250 mL Stericup-HV sterile vacuum filtration system (MilliPore) with a pore size of 0.45 µm. His-tagged BG505 SOSIP were purified using 5 mL bed volume of a nickel affinity HisTrap FF columns (GE Healthcare). Prior to loading the sample, the column was washed with 10 column volumes of washing buffer (50 mM $Na_2H_2CO_3$ PO_4, 300 mM NaCl) at pH 7. Imidazole was added to the filtered supernatant to a final concentration of 20 mM. Imidazole was used to minimize non-specific interactions. The sample was loaded onto the column at a speed of 1 mL/min. The column was washed with washing buffer (10 column volumes) and eluted in 5 column volumes of elution buffer (300 mM imidazole in washing buffer). The elution was concentrated by a Vivaspin column (50 kDa cut-off) to a volume of 1 mL.

To create an affinity column 1g of sepharose beads (enough to create a column that can bind 5-10 mg of protein) were added to 30ml of 1mM HCl and were incubated for 15 minutes

at room temperature with rotation before being spun down at 1000 rpm for 5 minutes with a high breaking setting on the centrifuge. The supernatant was then removed from the beads and another 30ml of 1mM HCl added to start the cycle again and this was repeated 4 times.

The purified ligand (PGT151) was then buffer exchanged with 0.1M Na_2HCO_3, 0.5 NaCl, pH 8 coupling buffer and resuspended to a final volume of 5ml. The solution containing the ligand was then added to the beads and rotated at room temperature for 2 hours. Prior to loading the beads to the column, the column was washed with 10ml of coupling buffer and emptied via gravity flow. The ligand bound to the beads was then added to the column and washed with 20ml of coupling buffer and the flow collected to analyse how much of the ligand has been coupled to the beads. To block any remaining binding sites 10ml of 0.1M Tris-HCl pH 8 blocking buffer was added, and the column closed and was left to stand for 2 hours at room temperature.

To prepare the column for use the blocking buffer is allowed to flow from the column before the column was washed with 10ml of alternating pH washing buffers, these were: 0.1M acetic acid/ sodium acetate, 0.5M NaCl, pH4 (washing buffer 1) followed by 0.1 M Tris-HCl, 0.5M NaCl, pH 8 (washing buffer 2). This was repeated 3 times before adding 10ml of 20mM Tris-HCl, 0.5M NaCl, pH 8 (washing buffer 3). At this point the column can be closed and stored at 4 degrees for later use.

The nickel elution was buffer exchanged from a phosphate buffer to a Tris buffer by spinning down the sample and then adding 2 ml of washing buffer 3. The washing buffer 3 in the column was then allowed to flow through the column before stopping the flow and slowly adding the Tris-buffered sample. The sample was then incubated for 5 minutes at room temperature to allow protein to bind to the ligand coupled beads. The column was then opened, and the flow collected in a 20ml Vivaspin column and the column was washed with 20ml of washing buffer 3, the flow was collected up to 20ml in the Vivaspin column. The column was then again stopped and 3ml of 3M MgCl was added to elute any bound proteins. After 5 minutes the column was opened and the elution collected into a VivaSpin column which had already had 5ml of 10mM Tris-HCl, 75mM NaCl, pH8 dialysing buffer added to prevent the protein from being exposed to high salt for too long. A further 2 ml of MgCl was added to the column to further elute any protein. Finally, 10ml of dialysing buffer was added to the column to make a final volume of 20ml in the VivaSpin column. Both the flow and the elution were then spun down at 4000 rpm for 15 minutes, with the elution then being spun down again with 20ml of dialysing buffer added to remove the high salt from the elution.

2.1.6 **In-gel digestion of N-glycans.**

The SDS-PAGE gel bands corresponding to gp120 and gp41 were excised and washed with 100% acetonitrile and water three times. The N-glycans for UPLC and ion mobility mass spectrometry analyses were then enzymatically released by within-gel digestion by PNGase F (New England BioLabs) for 16 h at 37°C.

2.1.7 **UPLC N-glycan analysis.**

Aliquots of released glycans were fluorescently labelled with procainamide using a prepared labelling solution (110 mg/mL procainamide, 60 mg/mL sodium cyanoborohydride in 30 % DMSO, 70 % acetic acid) at 65°C for 4 hours. Excess label and PNGaseF was removed using Spe-ed Amide 2 cartridges (Applied Separations). Glycans were analysed on a Waters Acquity H-Class UPLC instrument with a Glycan BEH Amide column (2.1 mm x 100 mm, 1.7 µM, Waters) and the following gradient: time (t) = 0: 22 % A, 78 % B (flow rate = 0.5 mL/min); t = 38.5: 44.1 % A, 55.9 % B (0.5 mL/min); t = 39.5: 100 % A, 0 % B (0.25 mL/min); t = 44.5: 100 % A, 0 % B (0.25 mL/min); t = 46.5: 22 % A, 78 % B (0.5 mL/min), where solvent A was 50 mM ammonium formate (pH 4.4) and B was acetonitrile. Fluorescence was measured at an excitation wavelength of 310 nm and an emission wavelength of 370 nm. Data were processed using Empower 3 software (Waters, Manchester, UK). The relative abundance of N-glycan structures was determined by digesting fluorescently labelled glycans with the following glycosidases (New England Biolabs) at 37°C for 16 h: Endoglycosidase H (Endo H), α2-3, 6, 8 neuraminidase and α2-3 neuraminidases. Glycans were extracted using a PVDF protein-binding membrane (Millipore) and analysed as described above.

2.1.8 **N-glycan mass spectrometry.**

Immediately prior to MS analysis, a separate unlabelled aliquot of glycans were further desalted for 30 min on top of a Nafion membrane and a trace amount of ammonium phosphate was added to promote phosphate adduct formation. Glycans were analysed by direct infusion using a Synapt G2Si instrument (Waters, Manchester, UK) with the following settings: capillary voltage, 0.8-1.0 kV; sample cone, 150 V; extraction cone, 150 V; cone gas, 40 l/h; source temperature, 80°C; trap collision voltage, 4-160 V; transfer collision voltage, 4 V; trap DC bias, 60 V; IMS wave velocity, 450 m/s; IMS wave height, 40 V; trap gas flow, 2 ml/min; IMS gas flow, 80 ml/min. Data were acquired and processed with MassLynx v4.1 and Driftscope version 2.8 software (Waters, Manchester, UK). Structural assignments were based on the identification of cross-ring and D-type fragments that are

characteristic for negatively charged glycan ions [221]. The nomenclature used to describe the fragment ions is that devised by Domon and Costello [127].

2.1.9 **Glycopeptide mass spectrometry.**

Env proteins were denatured, reduced, and alkylated by sequential 1 h incubations at room temperature (RT) in the following solutions: 50 mM Tris/HCl, pH 8.0 buffer containing 6 M urea and 5 mM dithiothreitol (DTT), followed by the addition of 20 mM iodoacetamide (IAA) for a further 1h at RT in the dark, and then additional DTT (20 mM), to eliminate any residual IAA. The alkylated trimers were buffer exchanged into 50 mM Tris/HCl, pH 8.0 using Vivaspin columns and digested separately with trypsin and chymotrypsin (Mass Spectrometry Grade, Promega) at a ratio of 1:30 (w/w). Reaction mixtures were dried, and glycopeptides were extracted using C18 Zip-tip (MerckMilipore) following the manufacturer's protocol.

Eluted glycopeptides were dried again and re-suspended in 0.1% formic acid prior to mass spectrometry analysis. An aliquot of intact glycopeptides was analysed by nanoLC-ESI MS with an Easy-nLC 1200 system coupled to a Fusion mass spectrometer (Thermo Fisher Scientific) using higher energy collisional dissociation (HCD) fragmentation. Peptides were separated using an EasySpray PepMap RSLC C18 column (75 µm x 75 cm) with a 275-minute linear gradient consisting of 0-32% acetonitrile in 0.1% formic acid over 240 minutes followed by 35 minutes of 80% acetonitrile in 0.1% formic acid. The flow rate was set to 300 nL/min. The spray voltage was set to 2.8 kV and the temperature of the heated capillary was set to 275 °C. HCD collision energy was set to 50%, appropriate for fragmentation of glycopeptide ions. Glycopeptide fragmentation data were extracted from the raw file using Byonic™ (Version 2.7) and Byologic™ software (Version 2.3; Protein Metrics Inc.). The glycopeptide fragmentation data were evaluated manually for each glycopeptide; the peptide was scored as true-positive when the correct b and y fragment ions were observed along with oxonium ions corresponding to the glycan identified. The chromatographic areas for each true-positive peptide with the same amino acid sequence were compared to allow the relative amounts of each glycoform at each site to be determined.

2.1.10 **Glycopeptide occupancy analysis.**

The remaining glycopeptides were first digested with Endo H (New England Biolabs) to deplete oligomannose- and hybrid-type glycans and leave a single GlcNAc residue at the corresponding site. The reaction mixture was then dried completely and resuspended in a mixture containing 50 mM ammonium bicarbonate and PNGase F (New England Biolabs) using only H_2O^{18} (Sigma-Aldrich) throughout. This second reaction cleaves the remaining complex-type glycans but leaves the GlcNAc residues remaining after Endo H cleavage

intact. The use of H_2O^{18} in this reaction enables complex glycan sites to be differentiated from unoccupied glycan sites as the hydrolysis of the glycosidic bond by PNGaseF leaves a heavy oxygen (O^{18}) isotope on the resulting aspartic acid residue. The resultant peptides were purified as outlined above and subjected to reverse-phase (RP) nanoLC-MS similar to the aforementioned glycopeptide analysis. Instead of the extensive N-glycan library used above, two modifications were searched for: +203 Da corresponding to a single GlcNAc, a residue of an oligomannose/hybrid glycan, and +3 Da corresponding to the O^{18} deamidation product of a complex glycan. A lower HCD energy of 27% was used as glycan fragmentation was not required. Data analysis was performed as above and the relative amounts of each glycan determined, including unoccupied peptides.

2.2 Materials and methods for Chapter 4

2.2.1 HEK293T cell culture (from the laboratory of Prof. Rogier Sanders, University of Amsterdam)

WT and NxT proteins, and the infectious virus stocks were prepared by transfecting 293T cells (American Type Culture Collection Cat. #11268) using the PEI MAX transfection reagent or Lipofectamine transfection reagent.

2.2.2 HEK293F cell culture and transfection (from the laboratory of Prof. Rogier Sanders, University of Amsterdam)

HEK293F cells were maintained at a density of 1-3x10^6 cells per mL at 37 degrees with 8% CO2 and 125rpm shaking. Plasmids encoding the WT, NxT, NxT T158S and NxT T135S T158S proteins containing a C-terminal D7324-tag were transiently co-transfected with a Furin-encoding plasmid (4:1) in HEK293F cells using PEI MAX transfection reagent. The cells were transfected at a density of 1x10^6 cells per ml and incubated for 6 days at 37 degrees with 8% CO2 and 125rpm shaking.

2.2.3 ExpiCHO-S cell culture and transfection (from the laboratory of Prof. John Moore, Weill Cornell Medical College)

ExpiCHO-S cells were maintained at a density of 1-2x10^6 cells per mL at 37 degrees with 8% CO2 and 135rpm shaking. The WT and NxT T35S T158S encoding plasmids containing a C-terminal D7324-tag were transiently co-transfected with a Furin-encoding plasmid (4:1) using FectoPRO transfection reagent as described by the manufacturer (VWR).

2.2.4 BG505 infectious molecular clones (from the laboratory of Prof. Rogier Sanders, University of Amsterdam)

The infectious molecular clone of LAI was used as the backbone for creating an infectious molecular clone containing BG505.T332N gp160. This clone contains a unique Sal1 restriction site 434 nucleotides upstream of the env start codon and a unique BamH1 site at the codons specifying amino acids G751 and S752 in LAI gp160 (HXB2 numbering). A DNA fragment containing the LAI sequences between the Sal1 site and the env start codon, followed by the BG505.T332N env sequences up to the BamH1 site, was synthesized (Genscript, Piscataway, NJ) and cloned into the LAI molecular clone backbone using Sal1 and BamH1. The resulting molecular clone encodes the complete BG505.T332N gp160 sequence, except for the C-terminal 106 amino acids of the cytoplasmic tail, which are derived from LAI gp160. The sequence was verified before the clone was used in virus infectivity and neutralization assays. The resulting virus was able to infect TZM-bl cells and replicate in PBMCs.

The BG505.T332N sequence was modified by mutating all PNGS to either NxT (NxT virus) or NxS (NxS virus). As described above, the resulting env sequences, between the Sal1 and BamH1 sites, where obtained from Genscript in a pUC57 cloning vector. The Sal1 and BamH1 BG505-NxT and BG505-NxS sequences were amplified by PCR, using In-Fusion cloning (Clontech), as described by the manufacturer. After the PCR, 0.5 µl DpnI (NEB) was added to each reaction and incubated for 1 h at 37 °C. Next, the PCR products were purified using a PCR clean-up kit (Macherey-Nagel) and the DNA concentrations were measured. The LAI env sequences between the Sal1 and BamH1 sites in the pLAI expression plasmid, were replaced by the BG505-NxT or BG505-NxS PCR fragments using the In-Fusion enzyme (Clontech) as described by the manufacturer. Plasmid DNA (5 µg) of the molecular clones was transfected into HEK293T cells to generate infectious virus stocks.

2.2.5 Env SOSIP trimer design and production (from the laboratory of Prof. Rogier Sanders, University of Amsterdam)

The HxB2 reference sequence has been used for the design of the HxB2 SOSIP trimer. The SOSIP trimer contains the v5 mutations and the 519S, 568D, 570H and 585H stabilizing mutations in gp41. The HxB2 NxT protein also has an PNGS added at position N411 to close a glycan hole. Env SOSIP trimer expression constructs in which all PNGS were mutated to either NxT (NxT protein) or NxS (NxS protein) were obtained from Genscript and cloned in the pPPI4 expression vector. The Env trimers were purified via PGT145-affinity chromatography and 2G12-affinity chromatography followed by SEC.

2.2.6 **Antibodies**

mAbs were obtained as gifts, or purchased, or expressed from plasmids, from the following sources directly or through the AIDS Reagents Reference Program: John Mascola and Peter Kwong (VRC01); Michel Nussenzweig (3BNC60, 3BC315); Dennis Burton (PG9, PG16, PGT121-123, PGT125-128, PGT130, PGT135, PGT145, PGDM1400 and PGT151); Barton Haynes (CH01); Polymun Scientific (2G12); Mark Connors (35O22); Ms C. Arnold (CA13 (ARP3119)), EU Programme EVA Centralized Facility for AIDS Reagents, NIBSC, UK (AVIP Contract Number LSHP-CT-2004-503487).

2.2.7 **Neutralization assays (from the laboratory of Prof. Rogier Sanders, University of Amsterdam)**

One day prior to infection, 1.7 X 10^4 TZM-bl cells per well were seeded on a 96-well plate in Dulbecco's Modified Eagles Medium (DMEM) containing 10% FCS, penicillin, and streptomycin (both at 100 U/ml) and incubated at 37°C in an atmosphere containing 5% CO2. A fixed amount of virus (2.5 ng/ml of p24-antigen equivalent) was incubated for 30 min at room temperature with serial 3-fold dilutions of each test mAb. This mixture was added to the cells and 40 µg/ml DEAE, in a total volume of 200 µl. Three days later, the medium was removed. The cells were washed once with PBS (150 mM NaCl, 50 mM sodium phosphate, pH 7.0) and lysed in Lysis Buffer, pH 7.8 (25 mM Glycylglycine (Gly-Gly), 15 mM MgSO4, 4 mM EGTA tetrasodium, 10% Triton-X). Luciferase activity was measured using a Bright-Glo kit (Promega, Madison, WI) and a Glomax Luminometer according to the manufacturer's instructions (Turner BioSystems, Sunnyvale, CA). All infection measurements were performed in quadruple. Uninfected cells were used to correct for background luciferase activity. The infectivity of each mutant without inhibitor was set at 100%. Nonlinear regression curves were determined, and 50% inhibitory concentrations (IC50) were calculated using a sigmoid function in Prism software version 8.

2.2.8 **SDS-PAGE and Blue Native-PAGE (from the laboratory of Prof. Rogier Sanders, University of Amsterdam)**

Env proteins were analysed using SDS-PAGE and BN-PAGE, blotted, and detected by using the CA13 (ARP3119) and 2G12 mAbs. In some cases, the gels were stained using Coomassie blue.

2.2.9 Negative stain electron microscopy (from the laboratory of Prof. Andrew Ward, The Scripps Research Institute)

The WT protein and NxT PNGS mutants were analysed by negative stain EM as previously described [218]. For the RM20E1 complexes, SOSIP trimers were incubated with 6-fold molar excess per protomer Fab at RT overnight. The complexes were purified using SEC on a Superose 6 Increase 10/300 GL (GE Healthcare) column.

2.2.10 Cryo electron microscopy (from the laboratory of Prof. Andrew Ward, The Scripps Research Institute)

The cryoEM dataset that led to EMDB-21232 was reprocessed with cryoSPARCv.1 to sort particles into classes based on RM20E1 Fab occupancy [218,222]. Final refinements were performed using cryoSPARCv.1.

2.2.11 Ni-NTA-capture ELISA (from the laboratory of Prof. Rogier Sanders, University of Amsterdam)

The Ni-NTA-capture ELISA has been described in detail elsewhere [76]. Purified His-tagged proteins were captured on HisSorb 96-well plates (Qiagen, Venlo, the Netherlands) and tested for Ab binding. Abs were detected with the goat-anti-human HRP-labelled Ab (SeraCare).

2.2.12 D7324-capture ELISA (from the laboratory of Prof. Rogier Sanders, University of Amsterdam)

Microlon 96-well half-area plates (Greiner Bio-One, Alphen aan den Rijn, the Netherlands) were coated with D7324 antibody (10 µg/ml: Aalto Bioreagents, Dublin, Ireland). Purified proteins (2.5 µg/ml) were subsequently captured on the ELISA plate wells and tested for Ab binding. Abs were detected with the goat-anti-human HRP-labelled Ab (SeraCare).

2.2.13 Biolayer interferometry (BLI) (from the laboratory of Prof. Rogier Sanders, University of Amsterdam)

Antibody binding to the PGT145-purified WT, NxT and NxT T135S T158S PNGS mutants was studied using a ForteBio Octet K2 instrument. All assays were performed at 30 °C with the agitation setting at 1000 rpm. Purified proteins and antibodies were diluted in running buffer (PBS, 0.1% BSA, 0.02% Tween 20) and analysed in a final volume of 250 µL/well. Antibody were loaded onto protein A sensors (ForteBio) at 2.0 µg/ml in running buffer until a binding threshold of 0.5 nM was reached. Trimer proteins were diluted in running buffer

at 40 nM or 600 nM for CH01, and association and dissociation were measured for 300 s. Trimer binding to a protein A sensor with no loaded antibody was measured to derive background values.

2.2.14 **Surface Plasmon Resonance (SPR) (from the laboratory of Prof. John Moore, Weill Cornell Medical College)**

NAb binding to the WT and NxT mutants was analysed by surface plasmon resonance (SPR). All experiments were performed at 25oC on a Biacore 3000 instrument (Cytiva, formerly GE Healthcare). Standard HBS-EP (0.01 M HEPES, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20, pH 7.4) was used as running buffer throughout the analysis.

Anti-histidine antibody coupled to CM5 (a-His-CM5) sensor-chip surfaces was used for capturing His-tagged the WT and NxT proteins. The anti-histidine antibody was covalently conjugated to CM5 sensor chips by amine coupling in accordance with the manufacturers' instructions, generating immobilization levels of ~104 RU (response units). The purified proteins were captured on the anti-His-CM5 sensor chips to a mean density of 530 RU (SD=9.3 RU). Fabs of PGT145 and PGT130 were titrated against His-tagged WT and NxT proteins, downwards from a concentration of 1µM, by two-fold dilution steps in running buffer, until no detectable signal was obtained. In each kinetic cycle, Fab binding was monitored for 300s of association, followed by 600s of dissociation. Maximum flow rate (50µl/min) was used to minimize mass-transfer limitation, which was absent according to kt analyses. At the end of each cycle, the sensor surface was regenerated by a single injection of glycine (10mM; pH 2.0) for 60s at a flow rate of 30µl/min. Titration data were analysed by fitting the models of the BIA-Evaluation software.

2.2.15 **N-glycan analysis using HILIC-UPLC**

N-linked glycans were released from gp140 in-gel using PNGase F (New England Biolabs). The released glycans were subsequently fluorescently labelled with procainamide using 110mg/ml procainamide and 60mg/ml sodium cyanoborohydride. Excess label and PNGase F were removed using Spe-ed Amide-2 cartridges (Applied Separations). Glycans were analysed on a Waters Acquity H-Class UPLC instrument with a Glycan BEH Amide column (2.1 mm x 100 mm, 1.7 µM, Waters). Fluorescence was measured, and data were processed using Empower 3 software (Waters, Manchester, UK). The relative abundance of oligomannose glycans was measured by digestion with Endoglycosidase H (Endo H; New England Biolabs). Digestion was performed overnight at 37 degrees. Digested glycans

were cleaned using a PVDF protein-binding membrane (Millipore) and analysed as described above.

2.2.16 **Site-specific glycan analysis using mass spectrometry**

Env proteins were denatured for 1h in 50 mM Tris/HCl, pH 8.0 containing 6M of urea and 5mM dithiothreitol (DTT). Next, the Env proteins were reduced and alkylated by adding 20 mM iodoacetamide (IAA) and incubated for 1h in the dark, followed by a 1h incubation with 20mM DTT to eliminate residual IAA. The alkylated Env proteins were buffer exchanged into 50 mM Tris/HCl, pH 8.0 using Vivaspin columns (3 kDa) and digested separately O/N using trypsin, chymotrypsin or subtilisin (Mass Spectrometry Grade, Promega) at a ratio of 1:30 (w/w). The next day, the peptides were dried and extracted using C18 Zip-tip (MerckMilipore). The peptides were dried again, re-suspended in 0.1% formic acid, and analysed by nanoLC-ESI MS with an Easy-nLC 1200 (Thermo Fisher Scientific) system coupled to a Fusion mass spectrometer (Thermo Fisher Scientific) using higher energy collision-induced dissociation (HCD) fragmentation. Peptides were separated using an EasySpray PepMap RSLC C18 column (75 µm x 75 cm). The LC conditions were as follows: 275-minute linear gradient consisting of 0-32% acetonitrile in 0.1% formic acid over 240 minutes followed by 35 minutes of 80% acetonitrile in 0.1% formic acid. The flow rate was set to 300 nL/min. The spray voltage was set to 2.7 kV and the temperature of the heated capillary was set to 40 °C. The ion transfer tube temperature was set to 275 °C. The scan range was 400-1600 *m/z*. The HCD collision energy was set to 50%, appropriate for fragmentation of glycopeptide ions. Precursor and fragment detection were performed using an Orbitrap at a resolution MS1= 100,000. MS2= 30,000. The AGC target for MS1=4e5 and MS2=5e4 and injection time: MS1=50ms MS2=54ms

Glycopeptide fragmentation data were extracted from the raw file using Byonic™ (Version 2.7) and Byologic™ software (Version 2.3; Protein Metrics Inc.). The glycopeptide fragmentation data were evaluated manually for each glycopeptide; the peptide was scored as true-positive when the correct b and y fragment ions were observed along with oxonium ions corresponding to the glycan identified. The MS data was searched using a standard library for HEK293F expressed BG505 SOSIP.664 (Seabright *et al.*, 2019). The precursor mass tolerance was set at 4ppm for MS1 and 10ppm for MS2. A 1% false discovery rate (FDR) was applied. The relative amounts of each glycan at each site as well as the unoccupied proportion were determined by comparing the extracted ion chromatographic areas for different glycopeptides with an identical peptide sequence. Glycans were categorized according to the composition detected. HexNAc(2)Hex(9−5) was classified as M9 to M5. HexNAc(3)Hex(5−6)X was classified as Hybrid with HexNAc(3)Fuc(1)X classified as Fhybrid. Complex-type glycans were classified according to the number of processed

antenna and fucosylation. If all of the following compositions have a fucose they are assigned into the FA categories. HexNAc(3)Hex(3-4)X is assigned as A1, HexNAc(4)X is A2/A1B, HexNAc(5)X is A3/A2B, and HexNAc(6)X is A4/A3B. As this fragmentation method does not provide linkage information isomers are grouped, so for example a triantennary glycan contains HexNAc (5) but so does a biantennary glycan with a bisect.

2.2.17 Site-specific analysis of low abundance N-glycan sites using mass spectrometry

To obtain data for sites that frequently present low intensity glycopeptide the glycans present on the glycopeptides were homogenized to boost the intensity of these peptides. This analysis loses fine processing information but enables the ratio of oligomannose: complex: unoccupied to be determined. The remaining glycopeptides were first digested with Endo H (New England Biolabs) to deplete oligomannose- and hybrid-type glycans and leave a single GlcNAc residue at the corresponding site. The reaction mixture was then dried completely and resuspended in a mixture containing 50 mM ammonium bicarbonate and PNGase F (New England Biolabs) using only H_2O^{18} (Sigma-Aldrich) throughout. This second reaction cleaves the remaining complex-type glycans but leaves the GlcNAc residues remaining after Endo H cleavage intact. The use of H2O18 in this reaction enables complex glycan sites to be differentiated from unoccupied glycan sites as the hydrolysis of the glycosidic bond by PNGaseF leaves a heavy oxygen isotope on the resulting aspartic acid residue. The resultant peptides were purified as outlined above and subjected to reverse-phase (RP) nanoLC-MS. Instead of the extensive N-glycan library used above, two modifications were searched for: +203 Da corresponding to a single GlcNAc, a remnant of an oligomannose/hybrid glycan, and +3 Da corresponding to the O18 deamidation product of a complex glycan. A lower HCD energy of 27% was used as glycan fragmentation was not required. Data analysis was performed as above and the relative amounts of each glycoform determined, including unoccupied peptides.

2.3 Materials and methods Chapter 5

2.3.1 Protein expression and purification

To express the prefusion S ectodomain, a gene encoding residues 1−1208 of SARS-CoV-2 S (GenBank: MN908947) with proline substitutions at residues 986 and 987, a “GSAS” substitution at the furin cleavage site (residues 682–685), a C-terminal T4 fibritin trimerization motif, an HRV3C protease cleavage site, a TwinStrepTag and an 8XHisTag

was synthesized and cloned into the mammalian expression vector pαH. This expression vector was used to transiently transfect FreeStyle293F cells (Thermo Fisher) using polyethylenimine. Protein was purified from filtered cell supernatants using StrepTactin resin (IBA) or nickel-affinity chromatography before being subjected to additional purification by size-exclusion chromatography using a Superose 6 10/300 column (GE Healthcare) in 2 mM Tris pH 8.0, 200 mM NaCl and 0.02% NaN_3.

2.3.2 Negative-stain electron microscopy and 2D class averaging (from the laboratory of Jason McLellan, University of Austin, Texas)

Purified SARS-CoV-2 spike was diluted to a concentration of 0.04 mg/mL using 2 mM Tris pH 8.0, 200 mM NaCl and 0.02% NaN_3 before being applied to a plasma cleaned CF400-Cu grid (Electron Microscopy Sciences). Protein was then stained using methylamine tungstate (Nanoprobes) before being allowed to dry at room temperature for 15 minutes. This grid was imaged in a Talos TEM (Thermo Fisher Scientific) equipped with a Ceta 16M detector. Micrographs were collected using TIA v4.14 software at a nominal magnification of 92,000×, corresponding to a calibrated pixel size of 1.63 Å/pix. CTF estimation, particle picking and 2D class averaging were performed using *cis*TEM [223].

2.3.3 Glycopeptide analysis by mass spectrometry

Three 30 µg aliquots of SARS-CoV-2 S protein, from three biological replicates, were denatured for 1h in 50 mM Tris/HCl, pH 8.0 containing 6 M of urea and 5 mM dithiothreitol (DTT). Next, the S protein were reduced and alkylated by adding 20 mM iodoacetamide (IAA) and incubated for 1h in the dark, followed by a 1h incubation with 20 mM DTT to eliminate residual IAA. The alkylated Env proteins were buffer exchanged into 50 mM Tris/HCl, pH 8.0 using Vivaspin columns (3 kDa) and digested separately overnight using trypsin chymotrypsin or alpha lytic protease (Mass Spectrometry Grade, Promega) at a ratio of 1:30 (w/w). The next day, the peptides were dried and extracted using C18 Zip-tip (MerckMilipore). The peptides were dried again, re-suspended in 0.1% formic acid, and analysed by nanoLC-ESI MS with an Easy-nLC 1200 (Thermo Fisher Scientific) system coupled to a Fusion mass spectrometer (Thermo Fisher Scientific) using higher energy collision-induced dissociation (HCD) fragmentation. Peptides were separated using an EasySpray PepMap RSLC C18 column (75 µm × 75 cm). A trapping column (PepMap 100 C18 3µM 75µM x 2cm) was used in line with the LC prior to separation with the analytical column. The LC conditions were as follows: 275-minute linear gradient consisting of 0-32% acetonitrile in 0.1% formic acid over 240 minutes followed by 35 minutes of 80% acetonitrile in 0.1% formic acid. The flow rate was set to 300 nL/min. The spray voltage was set to 2.7 kV and the temperature of the heated capillary was set to 40 °C. The ion transfer tube

temperature was set to 275 °C. The scan range was 400−1600 *m/z*. The HCD collision energy was set to 50%, appropriate for fragmentation of glycopeptide ions. Precursor and fragment detection were performed using an Orbitrap at a resolution MS^1= 100,000. MS^2= 30,000. The AGC target for MS^1=$4e^5$ and MS^2=$5e^4$ and injection time: MS^1=50ms MS^2=54ms.

Glycopeptide fragmentation data were extracted from the raw file using Byonic™ (Version 3.5) and Byologic™ software (Version 3.5; Protein Metrics Inc.). The glycopeptide fragmentation data were evaluated manually for each glycopeptide; the peptide was scored as true-positive when the correct b and y fragment ions were observed along with oxonium ions corresponding to the glycan identified. The MS data was searched using the Protein Metrics N-glycan library. The relative amounts of each glycan at each site as well as the unoccupied proportion were determined by comparing the extracted chromatographic areas for different glycotypes with an identical peptide sequence. All charge states for a single glycopeptide were summed. The precursor mass tolerance was set at 4 ppm and 10 ppm for fragments. A 1% false discovery rate (FDR) was applied. The relative amounts of each glycan at each site as well as the unoccupied proportion were determined by comparing the extracted ion chromatographic areas for different glycopeptides with an identical peptide sequence. Glycans were categorized according to the composition detected. HexNAc(2)Hex(9−5) was classified as M9 to M5. HexNAc(3)Hex(5−6)X was classified as Hybrid with HexNAc(3)Fuc(1)X classified as Fhybrid. Complex-type glycans were classified according to the number of processed antenna and fucosylation. If all of the following compositions have a fucose they are assigned into the FA categories. HexNAc(3)Hex(3-4)X is assigned as A1, HexNAc(4)X is A2/A1B, HexNAc(5)X is A3/A2B, and HexNAc(6)X is A4/A3B. As this fragmentation method does not provide linkage information compositional isomers are group, so for example a triantennary glycan contains HexNAc 5 but so does a biantennary glycans with a bisect. Any glycan containing at least one sialic acid was counted as sialylated. The quantifications of glycan compositions were represented as the mean of three biological replicates +/- standard error of the mean.

For O-linked glycan analysis, trypsin and alpha lytic protease-generated glycopeptides were treated with PNGase F prior to analysis to remove N-linked glycans. This was performed on a single biological replicate using an HCD energy of 27%. The MS data was searched using the Protein Metrics 70 common O-linked glycan library.

2.3.4 Model construction

Structural models of N-linked glycan presentation on SARS-CoV-2 were created using electron microscopy structures (PDB ID: 6VSB) along with complex-, hybrid-, and oligomannose-type N-linked glycans (PDB ID 4BYH, 4B7I, and 2WAH). A representative glycan presented at each site was modelled on to the N-linked carbohydrate attachment sites in Coot [224]

2.4 Materials and methods Chapter 6

2.4.1 SARS-CoV-2 S protein purification (from the laboratory of Prof. Bing Chen, Harvard University)

To express a stabilized ectodomain of Spike protein, a synthetic gene encoding residues 1−1208 of SARS-CoV-2 Spike with the furin cleavage site (residues 682–685) replaced by a “GGSG” sequence, proline substitutions at residues 986 and 987, and a foldon trimerization motif followed by a C-terminal 6xHisTag was created and cloned into the mammalian expression vector pCMV-IRES-puro (Codex BioSolutions, Inc, Gaithersburg, MD). The expression construct was transiently transfected in HEK 293T cells using polyethylenimine (Polysciences, Inc, Warrington, PA). Protein was purified from cell supernatants using Ni-NTA resin (Qiagen, Germany), the eluted fractions containing S protein were pooled, concentrated, and further purified by gel filtration chromatography on a Superose 6 column (GE Healthcare).

2.4.2 SARS-CoV-2 S protein purification (from the laboratory of Prof. David Stuart, University of Oxford)

The extracellular domain of SARS-CoV-2 spike was cloned into pHLsec vector[225] and encompassed residues M1-Q1208 (UniProtKB ID P0DTC2) with mutations R682G/R683S/R685S (furin recognition sequence), K986P/V987P, followed by Fibritin trimerization region, an HRV3C protease cleavage site, a His8-tag and a Twin-Strep-tag at the C-terminus [69,226]. Spike ectodomain was expressed by transient transfection HEK293T (ATCC, CRL-3216) cells for 6 days at 30 °C. Conditioned media was dialysed against 2x phosphate buffered saline pH 7.4 buffer (Sigma-Aldrich). The spike ectodomain was purified using Strep-Tactin Superflow resin (IBA Lifesciences) followed by size exclusion chromatography using Superose 6 Increase column (GE Healthcare) equilibrated in 200 mM NaCl, 2 mM Tris-HCl pH 8.0, 0.02% NaN_3 at 21 °C.

2.4.3 SARS-CoV-2 S protein purification (from the laboratory of Prof. Rogier Sanders, University of Amsterdam)

The SARS-CoV-2 S protein construct was designed and expressed as described previously [194]. Briefly, a SARS-CoV-2 S gene encoding residues 1-1138 (WuhanHu-1; GenBank: MN908947.3) was ordered (Genscript) and cloned into a pPPI4 plasmid containing a T4 trimerization domain followed by a hexahistidine tag by PstI-BamHI digestion and ligation. Modifications consist of substituting the amino acid positions 986 and 987 to prolines and removing the furin cleavage site by substituting amino acids 682-685 to glycine. S protein was expressed in HEK 293F cells. Approximately 1.0 million cells/ml, maintained in Freestyle medium (Gibco), were transfected by adding a mix of PEImax (1 µg/µl) and SARS-CoV-2 S plasmid (312.5 µg/l) in a 3:1 ratio in OptiMEM. After six days supernatants were centrifuged for 30 min at 4000 rpm and filtered using Steritop filter (Merck Millipore; 0.22 µm). S protein was purified from supernatant by affinity purification using Ni-NTA agarose beads and eluates were concentrated, and buffer exchanged to PBS using Vivaspin filters with a 100 kDa molecular weight cut-off (GE Healthcare). The S protein concentration was determined by the Nanodrop method using the proteins' peptidic molecular weight and extinction coefficient as determined by the online ExPASy software (ProtParam).

2.4.4 SARS-CoV-2 Spike Trimer protein production and purification from CHO cells (from Excellgene, Switzerland)

For CHO-based production of a trimeric Spike protein the construct Spike_ΔCter_ΔFurin_2P_T4_His was used [227]. Briefly, based on the CHO-vector pXLG-6 (ExcellGene SA), containing a puromycin resistance marker and optimized expression elements, the variant spike sequence was inserted, containing a scrambled furin cleavage site sequence, a two-proline sequence for blocking the protein in the prefusion form, a 3' T4-trimerization DNA, followed by a hexahistidine tag sequence. From chemical transfection (CHO4Tx®, ExcellGene) to clonal selection and isolation of protein producing cell lines, to subsequent expansion of cells and production – all steps were done under suspension culture in chemically defined media with strict adherence to regulatory recommendations.

Productions were done in orbitally shaken containers (TubeSpin bioreactor 600, TPP, Trasadingen, Switzerland) or in 5 L Erlenmeyer type shaker flasks in a Kuhner Shaker ISF4-X incubator (Adolf Kühner AG, Birsfelden, Switzerland), set to 37ºC, 5% CO2, humidity control and shaking at 150 rpm with a displacement radius of 50 mm. A fed-batch process

was implemented until harvest of supernatants on day 10. The viability of cell culture, reaching about 25 x 10^6 cells/mL remained in the 90-100% range until harvest. Production culture fluids were subjected to purification by affinity chromatography after depth filtration to remove cells. Loading onto, washing of and elution from a Ni-Sepharose column (Cytiva) were optimized, following the resin producers' suggestions. The eluted product stream was loaded on a Size-Exclusion column (SEC, Superdex 200pg, Cytiva) for further purification, following a concentration step through tangential flow filtration.

2.4.5 Plasmid construction and expression of Receptor Binding domain proteins (from the laboratory of Prof. Dennis Burton and Raiees Andrabi, The Scripps Research Institute)

To express the receptor-binding domains (RBDs) of SARS-CoV-2 (residues: 320-527), SARS-CoV-1 (residues: 307-513), and MERS-CoV (residues: 368- 587), DNA fragments were amplified by PCR using the SARS-CoV-2, SARS-CoV-1 and MERS-CoV S protein ectodomain plasmids as templates. The DNA fragments were cloned into the phCMV3 (Genlantis, USA) vector in frame with the Tissue Plasminogen Activator (TPA) leader sequence. To facilitate the purification and biotinylation of the protein, a C-terminal 6X HisTag, and an AviTag spaced by GS-linkers were added.

The RBD constructs were expressed by transient transfection of FreeStyle293F cells (Thermo Fisher). For a litre transfection, 350µg of the RBD-encoding plasmids was added to 16mL of Transfectagro (Corning) and 1.5mL of 40K PEI (1mg/mL) in 16mL of Transfectagro in separate 50mL conical tubes. The media containing the plasmids were filtered with a 0.22µm Steriflip Disposable Vaccum Filter Unit (Millipore Sigma) and gently mixed with PEI. The plasmid and PEI mixture was incubated for 30 minutes and gently added to 1L of cells at a concentration of 1million cells/mL. The culture supernatants were harvested after 4 days and loaded onto HisPur Ni-NTA resin column (Thermo Fisher). Following a wash with 25mM Imidazole, the protein from the column was eluted with 250mM Imidazole. The eluate was buffer exchanged with PBS and concentrated using 10K Amicon tuebs. The RBD proteins were further purified with size-exclusion chromatography using a Superdex 200 Increasse 10/300 GL column (GE Healthcare).

2.4.6 Sample preparation and analysis by LC-MS

Mass spectrometry data for the virion derived glycosylation was obtained using the protocols described in Yao *et al.* and reanalysed [228]. For recombinant S protein and RBD, 30 µg aliquots of SARS-CoV-2 S proteins were denatured for 1h in 50 mM Tris/HCl, pH 8.0 containing 6 M of urea and 5 mM of dithiothreitol (DTT). Next, the S proteins were reduced and alkylated by adding 20 mM IAA and incubated for 1hr in the dark, followed by incubation

with DTT to get rid of any residual IAA. The alkylated S proteins were buffer exchanged into 50 mM Tris/HCl, pH 8.0 using Vivaspin columns (3 kDa) and digested separately overnight using trypsin, chymotrypsin (Mass Spectrometry Grade, Promega) or alpha lytic protease (Sigma Aldrich) at a ratio of 1:30 (w/w). The peptides were dried and extracted using C18 Zip-tip (MerckMilipore). The peptides were dried again, re-suspended in 0.1% formic acid, and analysed by nanoLC-ESI MS with an Easy-nLC 1200 (Thermo Fisher Scientific) system coupled to an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) using higher energy collision-induced dissociation (HCD) fragmentation. Peptides were separated using an EasySpray PepMap RSLC C18 column (75 µm × 75 cm). A trapping column (PepMap 100 C18 3 µm (particle size), 75 µm × 2cm) was used in line with the LC prior to separation with the analytical column. The LC conditions were as follows: 275-minute linear gradient consisting of 0-32% acetonitrile in 0.1% formic acid over 240 minutes followed by 35 minutes of 80% acetonitrile in 0.1% formic acid. The flow rate was set to 300 nL/min. The spray voltage was set to 2.7 kV and the temperature of the heated capillary was set to 40 °C. The ion transfer tube temperature was set to 275 °C. The scan range was 400−1600 *m/z*. The HCD collision energy was set to 50%. Precursor and fragment detection were performed using Orbitrap at a resolution MS1 = 100,000. MS2 = 30,000. The AGC target for MS1 = $4e^5$ and MS2 = $5e^4$ and injection time: MS1 = 50 ms MS2 = 54 ms.

2.4.7 **Data processing of LC-MS data**

Glycopeptide fragmentation data were extracted from the raw file Byos (Version 3.5; Protein Metrics Inc.). The following parameters were used for data searches in Byonic: The precursor mass tolerance was set at 4 ppm and 10 ppm for fragments. Peptide modifications included in the search include: Cys carbamidomethyl, Met oxidation, Glu → pyroGlu, Gln → pyroGln and N deamidation. For each protease digest a separate search node was used with digestion parameters appropriate for each protease (Trypsin RK, Chymotrypsin YFW and ALP TASV) using a semi-specific search with 2 missed cleavages. A 1% false discovery rate (FDR) was applied. All three digests were combined into a single file for downstream analysis. All charge states for a single glycopeptide were summed. The glycopeptide fragmentation data were evaluated manually for each glycopeptide; the peptide was scored as true-positive when the correct b and y fragment ions were observed along with oxonium ions corresponding to the glycan identified. The protein metrics 309 mammalian N-glycan library was modified to include sulphated glycans and phosphorylated mannose species, although no phosphorylated mannose glycans were detected on any of the samples analysed. The relative amounts (determined by comparing the XIC of each glycopeptide, summing charge states) of each glycan at each site as well as the unoccupied proportion

were determined by comparing the extracted chromatographic areas for different glycotypes with an identical peptide sequence. Glycans were categorized according to the composition detected. HexNAc(2), Hex(9−3) was classified as M9 to M3. Any of these compositions that were detected with a fucose are classified as FM. HexNAc(3)Hex(5−6)Neu5Ac)(0-4) was classified as Hybrid with HexNAc(3)Hex(5-6)Fuc(1)NeuAc(0-1) classified as Fhybrid. Complex-type glycans were classified according to the number of processed antenna and fucosylation. Complex glycans are categorized as HexNAc(3)(X), HexNAc(3)(F)(X), HexNAc(4)(X), HexNAc(4)(F)(X), HexNAc(5)(X), HexNAc(5)(F)(X), HexNAc(6+)(X) and HexNAc(6+)(F)(X). Core glycans are any glycan smaller than HexNAc(2)Hex(3) Glycans were further classified in Figures 2 and 3 according to their processing state. A glycan was classified as fucosylated or sialylated if it contained at least one fucose or sialic acid residue. Agalactosylated glycans consisted of compositions in the above categories for complex-type glycans which possessed only 3 hexose residues, or if the composition was a hybrid-type glycan then this was increased to 5 hexoses. If the complex-type glycan did not contain sialic acid but more than 3/5 hexoses, then it was counted as galactosylated. Sulphated glycans were included in the above categories, so a glycan can be in both galactosylated and sulphated categories, for example.

2.4.8 Integrative modelling and molecular dynamics simulation (from the laboratory of Prof. Peter Bond, A*STAR Bioinformatics Institute, Singapore)

The model of S protein was built using Modeller 9.21 [229] with three structural templates: i) the cryo-EM structure of SARS-CoV-2 S ECD in the open state (PDB: 6VSB) [69], ii) the NMR structure of SAR-CoV S HR2 domain (PDB: 2FXP) [230], and iii) the NMR structure of HIV-1 gp-41 TM domain (PDB: 5JYN) [231]. Missing loops in the N-terminal domain and the C-terminal of the ECD were modelled using the cryo-EM structure of S ECD in the closed state, which was resolved at a higher resolution (PDB: 6XR8) [232]. A total of ten models were built and three top models (based on the lowest discrete optimized protein energy score [233] were subjected to further stereochemical assessment using Ramachandran analysis [234]. The model with the lowest number of outliers was then selected. Man-9 glycans were added to all 22 glycosylation sites using the CHARMM-GUI Glycan Reader and Modeller web server [235]. Three cysteine residues on the C-terminal domain (C1236, C1240 and C1243) were palmitoylated as they have been shown to be important in cell fusion [236]. The glycosylated S protein model was then embedded into a model ERGIC membrane built using CHARMM-GUI Membrane Builder [237], which contained 47% phosphatidylcholine (PC), 20% phosphatidylethanolamine (PE), 11% phosphatidylinositol phosphate (PIP), 7% phosphatidylserine (PS) and 15% cholesterol [238–240].

The system was parametrized using the CHARMM36 force field [241]. TIP3P water molecules were used to solvate the system and 0.15 M NaCl salt was added to neutralize it. The system was then subjected to stepwise energy minimization and equilibration with decreasing positional and dihedral restraints, following the standard CHARMM-GUI protocol [242]. A 200 ns production simulation was conducted with the temperature maintained at 310 K using the Nosé-Hoover thermostat [243,244]. The pressure was maintained at 1 atm using a semi-isotropic coupling to the Parrinello-Rahman barostat [245]. Electrostatics were calculated using the smooth particle mesh Ewald method [246] with a real-space cut-off of 1.2 nm, while the van der Waals interactions were truncated at 1.2 nm with a force switch smoothing function applied between 1.0-1.2 nm. The LINCS algorithm was utilized to constrain all covalent bonds involving hydrogen atoms [247] and a 2 fs integration time step was employed. All simulations were performed using GROMACS 2018 [248] and visualized in VMD [249]. ASA and glycan contact analyses were performed using the built-in GROMACS tools *gmx sasa* and *gmx select*.

2.5 Materials and methods Chapter 7

2.5.1 Expression and purification of trimeric SARS-CoV-2 spike

To express the prefusion S ectodomain, a gene encoding residues 1−1208 of SARS-CoV-2 S (GenBank: MN908947) with proline substitutions at residues 986 and 987, a "GSAS" substitution at the furin cleavage site (residues 682–685), a C-terminal T4 fibritin trimerization motif, an HRV3C protease cleavage site, a TwinStrepTag and an 8XHisTag was synthesized and cloned into the mammalian expression vector pαH. Expression plasmid encoding SARS-CoV-2 S glycoprotein was transiently transfected into Human Embryonic Kidney (HEK) 293F cells. Cells were maintained at a density of 0.2−3 $\times$ 10^6 cells per ml at 37°C, 8% CO_2 and 125 rpm shaking in FreeStyle 293F media (Fisher Scientific). Prior to transfection two solutions containing 25 mL Opti-MEM (Fisher Scientific) medium was prepared. Plasmid DNA was added to one to give a final concentration after transfection of 310 µg/L. Polyethylenimine (PEI) max reagent (1 mg/mL, pH 7) was added to the second solution to give a ratio of 3:1 PEI max: plasmid DNA. The two solutions were combined and incubated for 30 minutes at room temperature. Cells were transfected at a density of 1 $\times$ 10^6 cells per mL and incubated for 7 days at 37°C with 8% CO_2 and 125 rpm shaking.

After harvesting, the cells were spun down at 4000 rpm for 30 minutes and the supernatant applied to a 500 mL Stericup-HV sterile vacuum filtration system (Merck) with a pore size of 0.22 µm. The supernatant containing SARS-CoV-2 S protein was purified using 5 mL HisTrap FF column connected to an Akta Pure system (GE Healthcare). Prior to loading the sample, the column was washed with 10 column volumes of washing buffer (50 mM Na_2PO_4, 300 mM NaCl) at pH 7. The sample was loaded onto the column at a speed of 2 mL/min. The column was washed with washing buffer (10 column volumes) containing 50 mM imidazole and eluted in 3 column volumes of elution buffer (300 mM imidazole in washing buffer). The elution was concentrated by a Vivaspin column (100 kDa cut-off) to a volume of 1 mL and buffer exchanged to phosphate buffered saline (PBS).

The Superdex 200 16 600 column was washed with PBS at a rate of 1 mL/min. After 2 hours, 1 mL of the nickel affinity purified material was injected into the column. Fractions separated by SEC were pooled according to their corresponding peaks on the Size Exclusion chromatograms. The target fraction was concentrated in 100 kDa VivaSpin (GE healthcare) tubes to ~1 mL.

2.5.2 Expression and purification of ACE2

FreeStyle293F cells (Thermo Fisher) were transfected with polyethylenimine and a plasmid encoding residues 1−626 of human ACE2 with a C-terminal HRV3C protease cleavage site,

a Twin Strep Tag, and an 8×HisTag. This construct is identical to full length ACE2 except is truncated at position 626. This protein was expressed and purified identically as for the SARS-CoV2 glycoprotein, with the exception of a smaller Vivaspin cut-off being used for buffer exchanging. To produce oligomannose-type glycans on ACE2 20 μM kifunensine was added at the time of transfection. For ST6 co-expression DNA plasmid encoding ST6 was co-transfected with plasmid encoding ACE2 at a ratio of 2:1.

2.5.3 **Glycosidase digests**

To generate glycan modified ACE2 separate aliquots of ACE2 were digested with a range of glycosidases: α1-2,3,4,6 fucosidase, α2-3, 6, 8 sialidase and endoglycosidase H (NEB). Glycosidases were added at a ratio of 1:20 and incubated at 37°C overnight.

2.5.4 **His-Tag removal of ACE2**

Following purification, the His-Tag was removed from ACE2 using HRV3C protease cleavage (Thermo Fisher). Digestion was performed at a ratio of 1:20 HRV3C protease: ACE2 in 1 × HRV3C reaction buffer (Thermo Fisher) and incubated at 4°C overnight. To remove the HRV3C and uncleaved ACE2 nickel affinity chromatography was performed, except the flow through was collected rather than the elution.

2.5.5 **Mass spectrometry of glycopeptides**

30 μg aliquots of ACE2 and ACE2 glycan variants were denatured for 1h in 50 mM Tris/HCl, pH 8.0 containing 6 M of urea and 5 mM dithiothreitol (DTT). Next, the proteins were reduced and alkylated by adding 20 mM iodoacetamide (IAA) and incubated for 1h in the dark, followed by a 1h incubation with 20 mM DTT to eliminate residual IAA. The alkylated proteins were buffer exchanged into 50 mM Tris/HCl, pH 8.0 using Vivaspin columns (3 kDa) and digested separately overnight using trypsin, chymotrypsin, or alpha lytic protease (Mass Spectrometry Grade, Promega) at a ratio of 1:30 (w/w). The next day, the peptides were dried and extracted using C18 Zip-tip (MerckMilipore). The peptides were dried again, re-suspended in 0.1% formic acid, and analysed by nanoLC-ESI MS with an Easy-nLC 1200 (Thermo Fisher Scientific) system coupled to a Fusion mass spectrometer (Thermo Fisher Scientific) using higher energy collision-induced dissociation (HCD) fragmentation. Peptides were separated using an EasySpray PepMap RSLC C18 column (75 μm × 75 cm). A trapping column (PepMap 100 C18 3 μm (particle size), 75 μm × 2cm) was used in line with the LC prior to separation with the analytical column. The LC conditions were as follows: 275-minute linear gradient consisting of 0−32% acetonitrile in 0.1% formic acid over

240 minutes followed by 35 minutes of 80% acetonitrile in 0.1% formic acid. The flow rate was set to 300 nL/min. The spray voltage was set to 2.7 kV and the temperature of the heated capillary was set to 40 °C. The ion transfer tube temperature was set to 275 °C. The scan range was 400−1600 *m/z*. The HCD collision energy was set to 50%, appropriate for fragmentation of glycopeptide ions. Precursor and fragment detection were performed using an Orbitrap at a resolution MS^1= 100,000. MS^2= 30,000. The AGC target for MS^1=$4e^5$ and MS^2=$5e^4$ and injection time: MS^1=50 ms MS^2=54 ms.

Glycopeptide fragmentation data were extracted from the raw file using Byonic™ (Version 3.5) and Byologic™ software (Version 3.5; Protein Metrics Inc.). The glycopeptide fragmentation data were evaluated manually for each glycopeptide; the peptide was scored as true-positive when the correct b and y fragment ions were observed along with oxonium ions corresponding to the glycan identified. The MS data was searched using the Protein Metrics N-glycan library. The relative amounts of each glycan at each site as well as the unoccupied proportion were determined by comparing the extracted chromatographic areas for different glycotypes with an identical peptide sequence. All charge states for a single glycopeptide were summed. The precursor mass tolerance was set at 4 ppm and 10 ppm for fragments. A 1% false discovery rate (FDR) was applied. The relative amounts of each glycan at each site as well as the unoccupied proportion were determined by comparing the extracted ion chromatographic areas for different glycopeptides with an identical peptide sequence. Glycans were categorized according to the composition detected. HexNAc(2)Hex(9−5) was classified as M9 to M5. HexNAc(3)Hex(5−6)X was classified as Hybrid with HexNAc(3)Fuc(1)X classified as Fhybrid. Complex-type glycans were classified according to the number of processed antenna and fucosylation. If all of the following compositions have a fucose they are assigned into the FA categories. HexNAc(3)Hex(3-4)X is assigned as A1, HexNAc(4)X is A2/A1B, HexNAc(5)X is A3/A2B, and HexNAc(6)X is A4/A3B. As this fragmentation method does not provide linkage information compositional isomers are group, so for example a triantennary glycan contains HexNAc 5 but so does a biantennary glycans with a bisect. Any glycan containing at least one sialic acid was counted as sialylated.

2.5.6 **Surface plasmon resonance (SPR)**

SARS-COV-2 S and ACE2 proteins were buffer exchanged in HBS P+ buffer (Cytiva/GE Healthcare). All analysis was performed using a Biacore T200 (Cytiva/GE Healthcare). After removing contaminants via a pulse of EDTA (350 mM) for 1 min at a flow rate of 30 µL/min, the chip was loaded with Ni^{2+} by injecting $NiCl_2$ for 1 min at a flow rate of 10 µL/min. SARS-CoV2 S protein (50 µg/ml) was injected at 10 µL/min for 240s. Control channels were not

loaded with trimer. Control cycles were performed by flowing ACE2 over Ni^{2+}-loaded NTA in the absence of trimer. The analyte was injected into the trimer sample and control channels at a flow rate of 50 µL/min. Serial dilutions ranging from 200 nM to 3.125 nM were performed in triplicate along with HBS P+ buffer only as a control. Association was recorded for 300 s and dissociation for 600 s. After each cycle of interaction, the NTA-chip surface was regenerated with a pulse of EDTA (350 mM) for 1 min at a flow rate of 30 µL/min. A high flow rate of analyte solution (50 µL/min) was used to minimize mass-transport limitation. The resulting data were fit to a 1:1 binding model using Biacore Evaluation Software (GE Healthcare) and these fitted curves were used to calculate K_D. Paired t tests were used where possible to determine significance and for remaining samples where a direct experimental pairing was not possible an unpaired t test was performed. All statistics and graph plotting were performed using GraphPad prism v8.1.

Chapter 3 Site-Specific Glycosylation of Virion-Derived HIV-1 Env Is Mimicked by a Soluble Trimeric Immunogen.

3.1 Contributions

Chapter 3 involves a comparison of BG505 Env, produced recombinantly using the BG505 SOSIP.664 platform, and BG505 Env derived from infectious virus. External collaborators from the laboratory of Prof. Lifson (Frederick National Laboratory, USA) assisted this study by producing and purifying the viral Env, with the viral production outlined in **Appendix D.1.** Material for the analysis of GMP-grade BG505 SOSIP.664 was kindly provided by Dr Antu Dey of KBI Biopharma through the Bill & Melinda Gates Foundation [109]**.** In addition, Dr Gemma Seabright and Dr Weston Struwe (University of Oxford) collaborated on the glycan analytics, including sample preparation and data analysis. The sample manufacture was only possible due to the expertise in large scale viral production in Biosafety Level 3 facilities provided by the laboratory of and all contributors are thanked for their work.

3.2 Introduction

Fundamental features of retroviral biology hinder the development of an effective vaccine against human immunodeficiency virus-1 (HIV-1). A rapid mutation rate creates profound antigenic diversity that can only be countered by immune responses of exceptional breadth. A barrier to a neutralizing antibody (nAb)-based vaccine is the dense coat of glycans covering the surface of the envelope glycoprotein (Env) spike that restricts the immunogenicity of Env proteins and limits the binding of antibodies that are elicited.

The HIV-1 Env glycans were long considered to be merely a passive shield; the number of individual glycan sites and their precise locations on Env evolve during infection as a dynamic response to nAb selection pressures [250–252]. However, various broadly neutralizing antibodies (bnAbs) of substantial potency can be isolated from infected individuals that do not merely penetrate the glycan shield but also recognize elements of it. Passively administered bnAbs protect macaques or humanized mice against virus challenge with an efficiency that is generally predicted by *in vitro* neutralization titres [253,254]. The existence of bnAbs, combined with an understanding of their properties, together underpin vaccine designs intended to induce similar antibodies. Generally, these concepts involve producing recombinant mimics of the native, virion associated Env spike that present multiple bnAb epitopes and/or their predicted human germline precursors [66,255].

The most commonly used HIV-1 Env immunogen design platform is based on SOSIP.664 trimers (SOSIP), recombinant proteins rendered soluble by truncation at position 664 and engineered for improved stability by the addition of a disulphide bond ('SOS') and an isoleucine to proline mutation ('IP') [66,93,219]. The prototype, and most widely studied, SOSIP trimer is derived from the BG505 clade A paediatric isolate [256]. BG505 SOSIP.664 trimers have Env spike-mimicking antigenicity and structural properties [76,109]. They have now been produced in gram-quantities in Chinese hamster ovary (CHO) cells under current Good Manufacturing Practice (cGMP) conditions for human trials [109]. As glycans both shield and form bnAb epitopes [84], and that viral neutralization by bnAbs is influenced by glycan heterogeneity [257], it is of substantial relevance to characterise them on both the viral target and HIV-1 Env immunogens. The unusually high surface density of glycans on these proteins restricts the extent to which multiple individual sites are processed within the ER and the Golgi apparatus [105]. Env processing is restricted on two levels. Thus, monomeric gp120 subunits and gp140 uncleaved pseudotrimers that are predominantly in non-native configurations all carry a range of highly processed, complex-type glycans together with a smaller population of unprocessed oligomannose-type glycans ($Man_{5-9}GlcNAc_2$, hereafter referred to as Man5-9) [105,258,259]. Additional influences were revealed when SOSIP trimers were found to bear substantially increased levels of oligomannose-type glycans [100,103,118,162]. The packing of gp120 subunits into native-like trimers further shapes Env glycosylation because key glycan processing enzymes have a limited ability to encounter their substrates; access is sterically hindered by nearby glycan and protein elements [84,118]. However, the complex-type glycans found at multiple sites on Env proteins are not constrained in the same way as oligomannose sites but are instead influenced by the prevailing secretory environment in a cell-specific manner [260]. This factor is relevant to immunogen design as certain complex-type glycans can also contribute to bnAb epitopes and, perhaps, to immunogenicity [82,84,180,261,262].

Overall, site-specific analyses have revealed a conservation in the distribution of mannose moieties across the surface of SOSIP trimers [100,162]. Regions of oligomannose-type glycans located at equivalent positions on the gp120 monomer and SOSIP trimer constitute the 'intrinsic mannose patch' (IMP) [118,162], whereas other such glycans that are trimer-specific are designated the 'trimer-associated mannose patch' (TAMP) [118,175]. Analysis of cell-line derived HIV-1 BaL virions [163] and membrane-tethered forms of Env have suggested there could be factors leading to divergent glycosylation compared to soluble trimers [263].

Here, the glycosylation of Env derived from the BG505.T332N-LAI virus, produced in a human lymphoid cell line, with the sequenced-matched BG505 SOSIP.664 soluble trimers expressed in CHO or 293F cells, is compared. The abundance of oligomannose and

complex-type N-linked glycans on the gp120 and gp41 subunits was determined by chromatographic methods and the precise glycan structures defined using by ion-mobility tandem mass spectrometry (IM-MS/MS). To characterize site-specific differences, complimentary approaches involving the comparison of intact and glycosidase-treated glycopeptides were used. Sufficient gp120, but not gp41, could be isolated from virions for glycopeptide analysis. Overall, this information on the glycan-dependent or glycan-influenced bnAb epitopes present on infectious HIV-1 virions will facilitate the continued development of Env immunization strategies intended to eventually elicit bnAb-like antibodies.

3.3 Preparation of HIV-1 virion-derived Env

An infectious stock of HIV-1 BG505.T332N-LAI was derived by transfecting 293T cells with the infectious molecular clone (IMC) (**Appendix D.1**). The IMC Env glycoproteins have gp120 and gp41-ectodomain subunits derived from the primary R5 Tier-2, clade A BG505 week 6 isolate, but with the N332-glycan knocked-in to mimic the BG505 SOSIP.664 soluble trimer sequence [76,193]. The transfection stock was used to infect the CD4+, CCR5+ A66-R5 T cell line to provide the large amount of virus required for glycan analysis **(Appendix D.1** and **D.5**).

A 1.9 mL aliquot of the 1000-fold concentrated virus stock was processed by high performance liquid chromatography (HPLC) (**Appendix D.1C**). The fractions containing gp120 and gp41 subunits were then run on an SDS-PAGE gel (**Appendix D.1D** and **D.1E**). The gp120- and gp41-containing bands were excised from the gel for analysis of released glycans by ultra-high-performance liquid chromatography (UPLC) and mass spectrometry. A separate 2.7 mL aliquot of the virus stock was then purified (**Appendix D.1F and D.1G**). The pooled gp120 fractions (~70 µg) were subsequently utilized for site-specific glycan analyses.

3.4 Relative abundance of oligomannose- vs. complex-type glycans

The purified gp120 and gp41 gel bands shown in **Appendix D.1D** and **D.1E** were subjected to in-gel PNGase F digestion, which removes both oligomannose- and complex-type glycans. The relative amounts of oligomannose- vs. complex-type glycans on the virion gp120 and gp41 subunits were compared to recombinant BG505 SOSIP.664 trimers that were produced either transiently in 293F cells and purified via the PGT151 bnAb or under cGMP conditions in a stable CHO cell line and purified via 2G12/SEC [109,260]. These comparisons are relevant, because initial characterizations of the BG505 trimer glycans

used 293F cell-produced proteins, while human clinical studies will be performed with cGMP-grade trimers derived from CHO cells [109]. Furthermore, the goal of the clinical studies is to elicit nAbs that recognize functional Env proteins present on infectious virions that are produced predominantly in human lymphoid cells.

Differences between the Env forms were quantified using an UPLC-based analytical procedure whereby differential sensitivity to endoglycosidase H (Endo H) reveals the abundances of oligomannose- and hybrid-type glycans (**Figure 3-1**; green chromatograms), and complex-type glycans (**Figure 3-1**; magenta chromatograms). The N-glycans on the virion-derived gp120 proteins were 50% oligomannose-type and 50% complex-type (**Figure 3-1A**), whereas the gp120 subunits of the SOSIP.664 trimers were enriched for oligomannose-type structures: 63% for the 293F cell product (**Figure 3-1B**) and 73% for CHO (**Figure 3-1C**) [109]. There were also differences in the oligomannose-type structures present on the three sources of gp120 (**Figure 3-1** and **Appendix D.6**). For example, Man9 and Man8 moieties were prominent in the SOSIP.664 gp120 chromatograms. Similarly, virion gp41 contained a lower overall content of oligomannose-type N-glycans (4%, **Figure 3-1D**) compared to gp41 from SOSIP.664 trimers (22% (293F) and 26% (CHO), **Figure 3-1E** and **Figure 3-1F** respectively). Taken together, the glycans on the viral Env proteins are somewhat more processed than those on the recombinant trimers.

Larger and more elaborated, complex-type glycans were also relatively abundant on the virus-derived gp120 and gp41 subunits, compared to SOSIP.664 (**Figure 3-1**; magenta chromatograms). This is exemplified by N-glycans containing either α2-3- or α2-6-linked terminal sialic acids. Overall, the extent of gp120 sialylation was higher for BG505.T332N-LAI virions (16%) than for 293F-derived BG505 SOSIP.664 trimers (5%), while for the gp41 subunits the corresponding values were 26% and 8%, respectively. The ratio of α2-6/α2-3 sialic acids on the gp41 subunits was 1.8 for virions and 0.8 for 293F-derived trimers (**Appendix D.6**), a finding consistent with earlier observations that α2-6-linked sialic acid moieties were prevalent on peripheral blood mononuclear cells (PBMC)-derived virion gp41 [103]. However, compared to gp41, the gp120 proteins from all three Env proteins contained a smaller population of sialic acid-containing glycans (**Appendix D.6**). This observation may reflect a generally lower efficiency of glycan processing on gp120 subunits than on gp41, irrespective of the Env protein source.

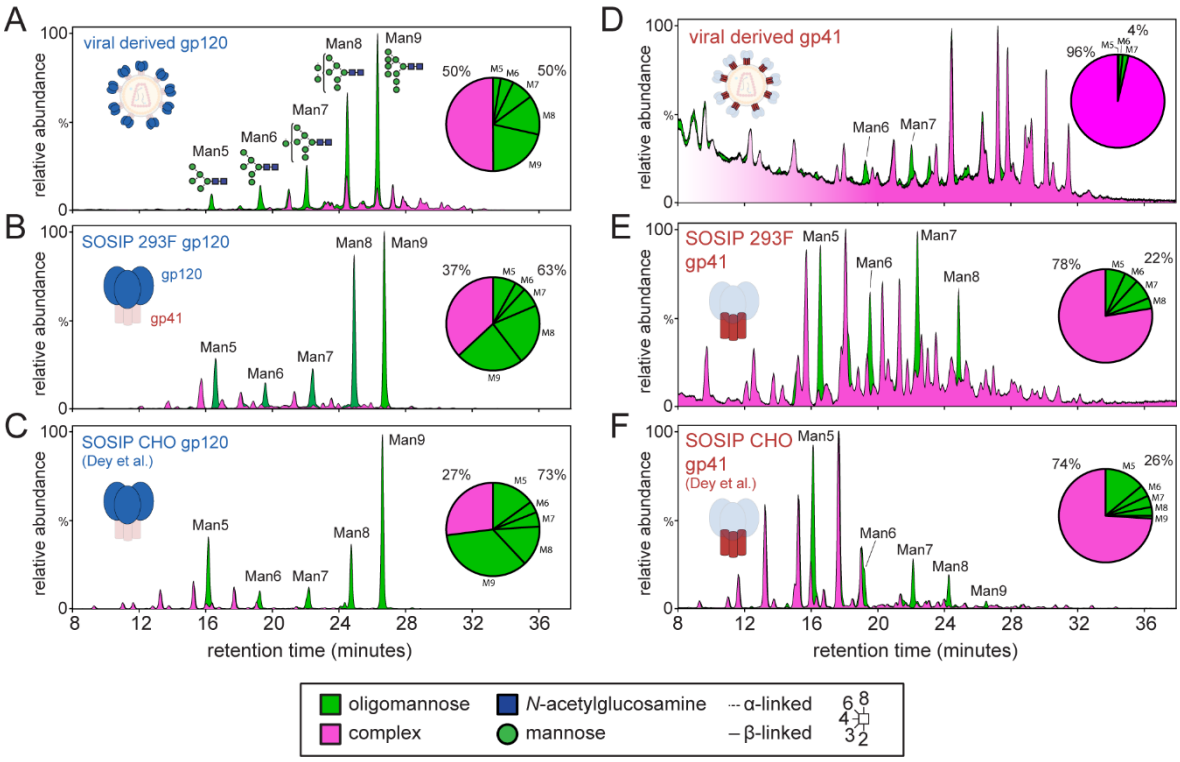


Figure 3-1: UPLC analysis of N-glycans from gp120 and gp41.
(A) UPLC chromatogram of N-glycans from BG505.T332N-LAI gp120 gel bands. Peaks sensitive to endoglycosidase H (endoH) digestion are coloured green and represent oligomannose-type glycans. Peaks resistant to endoH are coloured magenta and represent complex-type glycans. Pie charts depict the quantification of oligomannose and complex-type glycans in each sample. Peaks corresponding to oligomannose-type glycans are annotated with the number of mannose residues (Man5-9). (B) UPLC chromatogram of N-glycans released from gp120 from BG505 SOSIP.664 trimers produced in 293F cells. (C) Reproduction of the UPLC analysis of N-glycans released from BG505 SOSIP.664 trimers produced in CHO cells [109]. Panels D, E and F show data for gp41 and correspond to the gp120 panels A, B and C, respectively.

3.5 Analysis of glycans by ion-mobility mass spectrometry

To guide subsequent glycopeptide analyses, ion-mobility MS on a separate, unlabelled aliquot of glycans released from gp120 and gp41 subunits was performed. Comparing the BG505.T332N-LAI virion Env and the 293F-derived SOSIP.664 trimers allow fine structure determinations, namely glycan linkage and branching and the compositions of individual glycans. In this way, extensive sample-specific glycan libraries were created that were then used to search the corresponding site-specific data (**Figure 3-2**). In addition, collision-induced dissociation (CID) fragmentation in negative mode was used to obtain diagnostic fragment ions of the fine structures of the detected glycans (**Appendix D.2, D.3 and D.4**). The oligomannose-type structures on gp120 from the BG505.T332N-LAI virus (**Figure 3-2A**) and 293F cell-derived BG505 SOSIP.664 trimers (**Figure D.2 and D.3**) were highly similar. The major non-oligomannose-type glycans on BG505.T332N-LAI virion gp120 consisted of various hybrid-type structures, i.e., glycans on which some arms have a

terminal mannose while others are more extensively processed. (**Figure 3-2**, **Appendix D.4A**). Consistent with the overall trend detected by UPLC (**Figure 3-1**), the ion intensities of singly and doubly charged ions corresponding to more elaborated complex-type glycans were significantly higher in the virion-derived material compared to that of SOSIP.664 trimers (**Figure 3-2**).

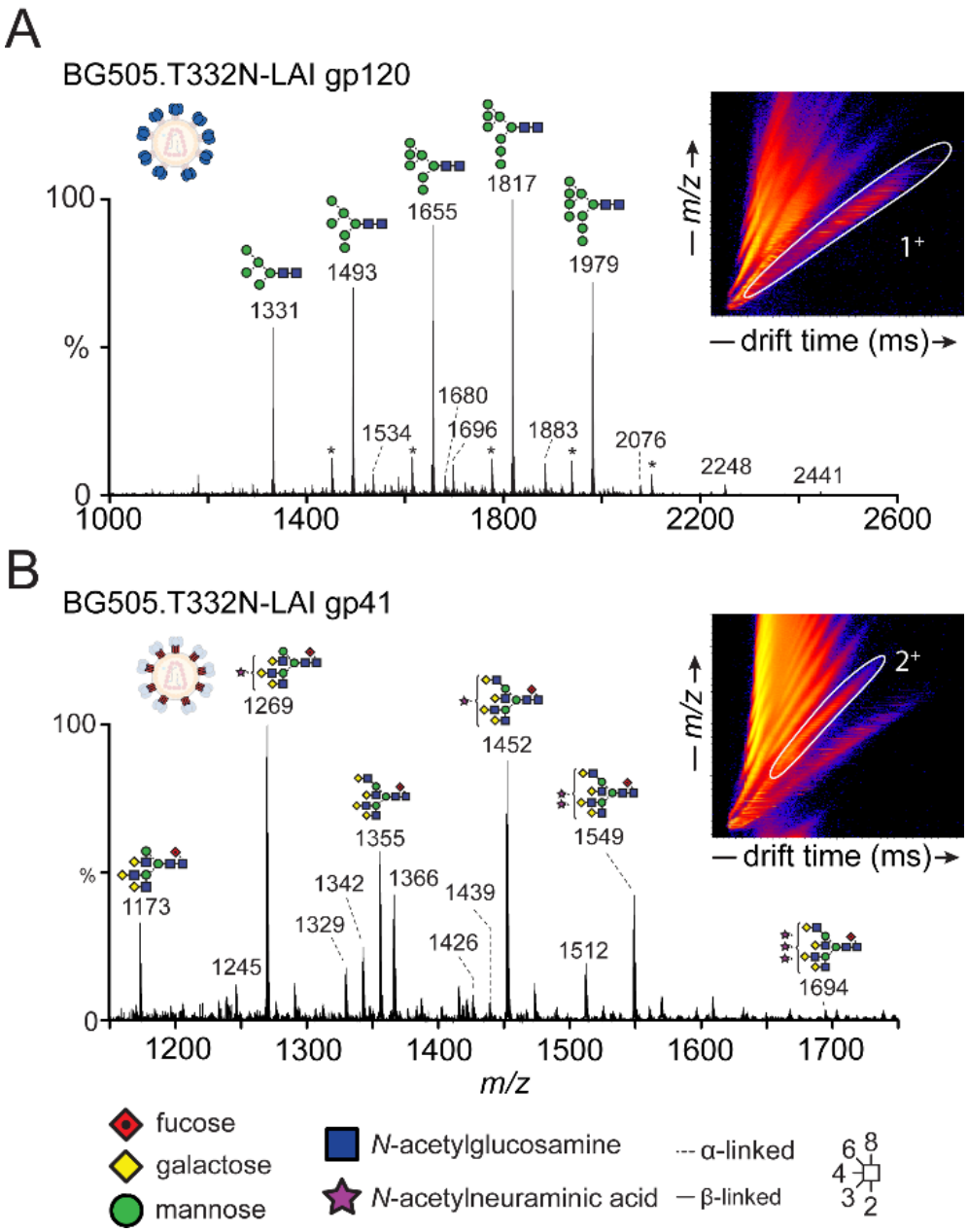


Figure 3-2 Ion mobility mass spectrometric analysis of BG505.T332N-LAI virion derived gp120 and gp41 N-glycans.
(A) Mobility-extracted singly charged negative ions from BG505.T332N-LAI gp120. The corresponding singly charged ions (1+) are encircled in white in the ion mobility drift plot (inset). A series of carbohydrate ions (marked with *) corresponding to a loss of a single N-acetyl group (-42 Da) from $Man_6GlcNAc_2$-$Glc_1Man_9GlcNAc_2$ compositions were visible in the spectrum of virion gp120. These peaks have been defined as contaminants as the N-glycan core consists of 2 GlcNAc residues (B) Mobility-extracted doubly charged negative ions from BG505.T332N-LAI gp41. The corresponding doubly charged ions (2+) are encircled in white in the ion mobility drift plot.

3.6 Complex-type structures on gp41

The glycans present on BG505.T332N-LAI virion gp41 were primarily core fucosylated, tri- and tetra-antennary species containing a variable number of terminal sialic acid residues (**Figure 3-2B**). Consistent with the UPLC analysis, larger complex-type glycans were more abundant on gp41 from virions than from 293F cell-derived trimers (**Appendix D.3B**), although generally similar structures were present in both cases. One caveat is that the producer cell is different for the BG505.T332N-LAI virion (A66-R5 lymphoid cells) and BG505 SOSIP.664 (CHO or 293F cells) Env proteins. Even so, membrane-localization does not lead to a sterically driven increase in oligomannose-type glycans. Membrane tethering may, however, influence the degree of branching and elaboration of complex-type glycans by driving proximity to processing enzymes and influencing the kinetics of egress through the secretory system.

3.7 Quantifying the amounts of under processed glycans across gp120

Understanding how HIV-1 Env proteins are glycosylated and defining what glycoforms are present is important as it is now thought that the entire surface of the trimer contributes to bNAb epitopes [84,85]. The methodology paradigm for defining which type of glycan (i.e., complex *vs.* oligomannose) is present at each potential glycosylation site involves a proteomic-based analysis of glycopeptides that have been enzymatically treated to create mass modifications that are specific to each glycan type [162]. This method has proved highly robust for determining the ratio of oligomannose-type *vs.* complex glycans, even when only low amounts of sample are available [264]. As such this method was applied to study the gp120 subunits of the BG505.T332N-LAI virus.

The proportion of oligomannose-type, complex-type, and unoccupied potential N-glycosylation sites (PNGs) across all 24 sites on the BG505 virion gp120 was determined using this method. However, insufficient gp41 was isolated for a similar analysis of that subunit. For comparison, the analysis was repeated using recombinant BG505 SOSIP.664 trimers that were produced either transiently in 293F cells and purified via the PGT151 bnAb or in a stable CHO cell line and purified via 2G12/SEC. These comparisons are relevant, as initial characterizations of the BG505 trimer glycans used 293F cell-produced proteins, while human clinical studies will be performed with cGMP-grade trimers derived from CHO cells [109]. Furthermore, the goal of the clinical studies is to elicit nAbs that recognize functional Env proteins present on infectious virions.

The site-specific analysis revealed that the processing of N glycans is conserved at certain sites and divergent at other sites on gp120s from the BG505.T332N-LAI virus

compared to both the 293F-cell and the CHO-cell derived SOSIP.664 trimers. The elevation in processing was not universal in that it was localized towards the apex of the trimer at sites such as N133, N137, and N197 (**Figure 3-3**). Hence, the previously described mannose patches (IMP and TAMP) are a feature shared not just by the two batches of recombinant SOSIP.664 trimer (293F and CHO) but also by the BG505.T332N-LAI virus [103,118]. The N332-glycan has been designated a 'supersite of vulnerability' in that it is a key component of multiple bnAbs [179,265]. This glycan is consistently in high oligomannose form; the recombinant trimers therefore correctly present the glycan components of the epitopes for bnAbs that recognize the mannose patch. This conclusion is exemplified by the recent characterization of BG18, a highly potent bnAb that binds BG505 SOSIP.664 trimers via critical contacts with the N332-glycan [266]. The processing of the previously characterized TAMP was also conserved across the three trimer preparations. Thus, sites such as N160 and N156, which present complex-type glycans when BG505 gp120 is expressed as a monomer, are predominantly in oligomannose form on the virion-derived trimers and their recombinant SOSIP.664 comparators. This finding again emphasizes the important influence of trimer formation on Env glycosylation and shows that it applies also to virion-derived proteins and not just recombinant trimers [79,162].

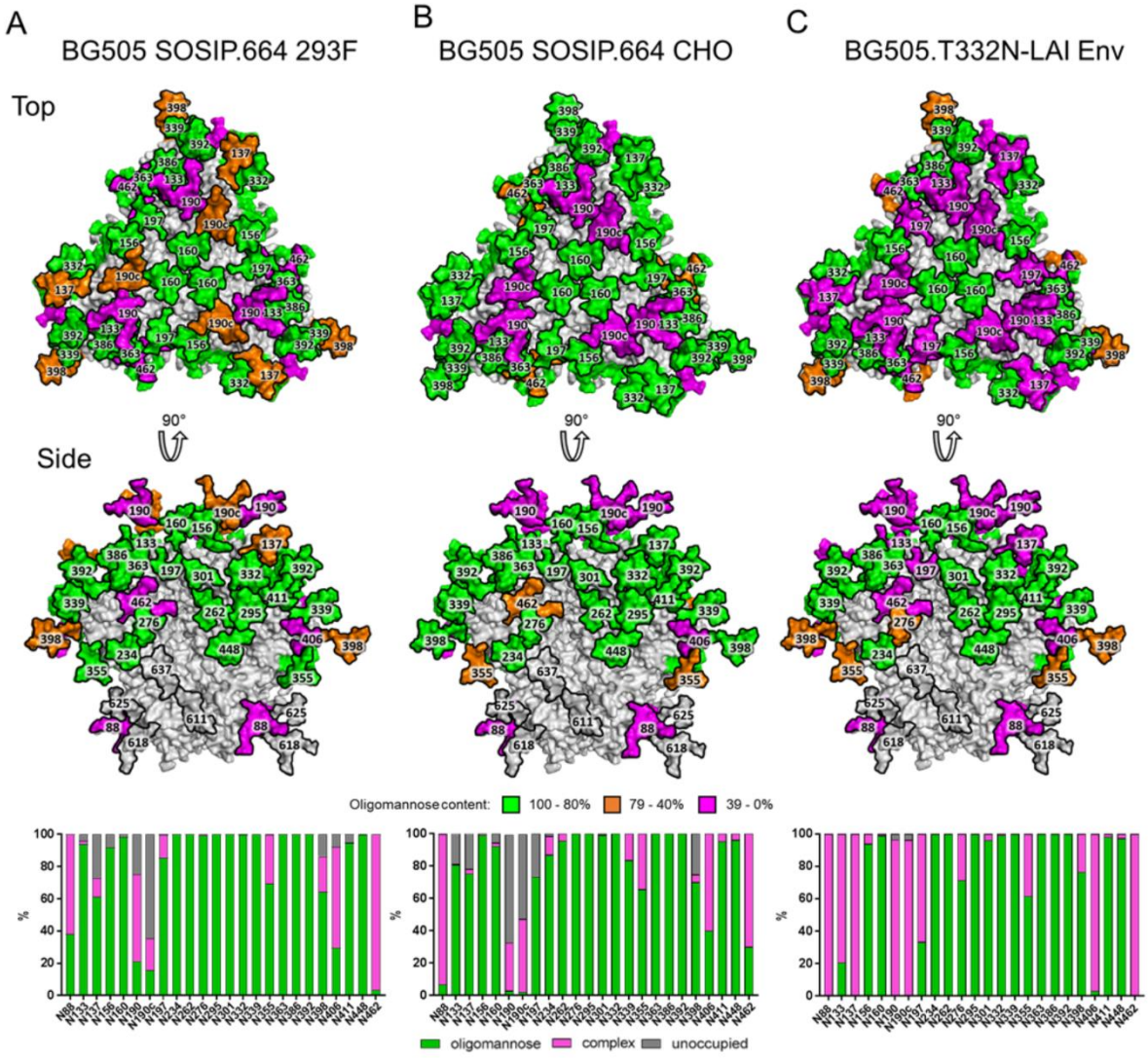


Figure 3-3 Determination of the predominant glycoforms presented at each site on soluble BG505 SOSIP.664 gp120 expressed in (A) 293F cells, (B) CHO cells and from (C) BG505.T332N-LAI virus.
The predominant glycoform at each site, as determined using glycosidase-treated glycopeptides, is modelled onto the cryo-EM model of BG505 SOSIP.664 [100]. The bar graphs represent the relative amounts of digested glycopeptides possessing the footprints for oligomannose glycans (green), complex-type glycans (magenta) and unoccupied PNGs (grey) at each gp120 site, listed from N- to C-terminus.

One notable difference was that the occupancy of glycan sites at certain regions across gp120 was lower for the soluble, truncated, recombinant BG505 SOSIP.664 trimers than for their membrane-associated, full-length, virion-derived counterpart. This outcome was most noticeable around the V1/V2 region, where the N137, N190 and N190c sites were occupied by N-linked glycans on only ~30-80% of the gp120 populations derived from the 293F and CHO cell-derived BG505 SOSIP.664 trimers (**Figure 3-3**). In contrast, the glycan-occupancy levels of the same gp120 sites on the BG505.T332N-LAI virion gp120 were >95% (**Figure 3-3C**). The codon-optimization process that is used to maximize the production of recombinant proteins may result in faster early folding events during Env biosynthesis and therefore reduce the likelihood of an N-glycan being attached. Such an

effect may be particularly apparent at sites such as the V1/V2 region that contain several densely packed glycan sites. While over expression levels do not appear to have significant impact on glycan processing [109], evidence is emerging that glycan processing may be influenced by choice of signal peptide [267].

3.8 Intact glycopeptide bnAb epitope mapping on virion Env

Although glycosidase-treated peptides provide information about the occupancy state of specific glycan sites using only small amounts of sample, the enzymatic release of glycans to generate the oligomannose/complex-type glycan footprints do result in the loss of information regarding the precise identity of glycan structures. Thus, it is useful to know the fine processing states of N-linked glycans to ensure that trimer immunogens correctly reproduce key bnAb epitopes at a site-specific level. To characterize the site-specific glycans at such sites on BG505.T332N-LAI virion gp120 subunits, a glycopeptide-based LC-MS approach was employed, which had been previously validated [100]. This analysis was performed on the same batch of material as that described for the glycosidase digested peptides. The inherent heterogeneity of N-linked glycosylation results in a diverse range of glycopeptides when analysed by LC-MS. To simplify the categorization process, from here on glycans are grouped according to the number of mannose residues they possess (for oligomannose-type), the number of fully processed antenna and the presence or absence of fucose (complex-type). By determining the fine processing states of glycan epitopes across viral Env and pattern-matching the outcome to neutralization data derived using bnAbs that recognize the corresponding sites, it was demonstrated that the processing of critical N-glycans was conserved between virion-derived and recombinant trimers. Here, new glycopeptide mass spectrometry data for the BG505.T332N-LAI virion and 293F cell-derived BG505 SOSIP.664 trimers were compared with previously published information on the CHO cell-derived, clinical grade BG505 SOSIP.664 trimers [109]. The commonality of the procedures used justifies this cross-study comparison.

A major class of bnAbs is directed against oligomannose-type glycans across the outer domain of gp120. Glycopeptide analysis revealed the fine processing states for several N-linked glycan sites recognized by such bnAbs on the virion-derived trimer (**Figure 3-4A**). Comparing the three versions of the BG505 trimer reveals the overall integrity of oligomannose processing, with $Man_9GlcNAc_2$ dominating the majority of sites across the IMP. This finding applies to N332 and N339, which are critical to the 2G12 bnAb epitope [268], and is therefore consistent with the ability of 2G12 to neutralize the BG505.T332N-LAI virus (**Figure 3-4C** and **F**). The only notable deviation in the fine processing of oligomannose-type glycans was seen at N363, where $Man_8GlcNAc_2$ predominated on

virion-derived trimers but $Man_9GlcNAc_2$ did so on both of the recombinant trimer preparations.

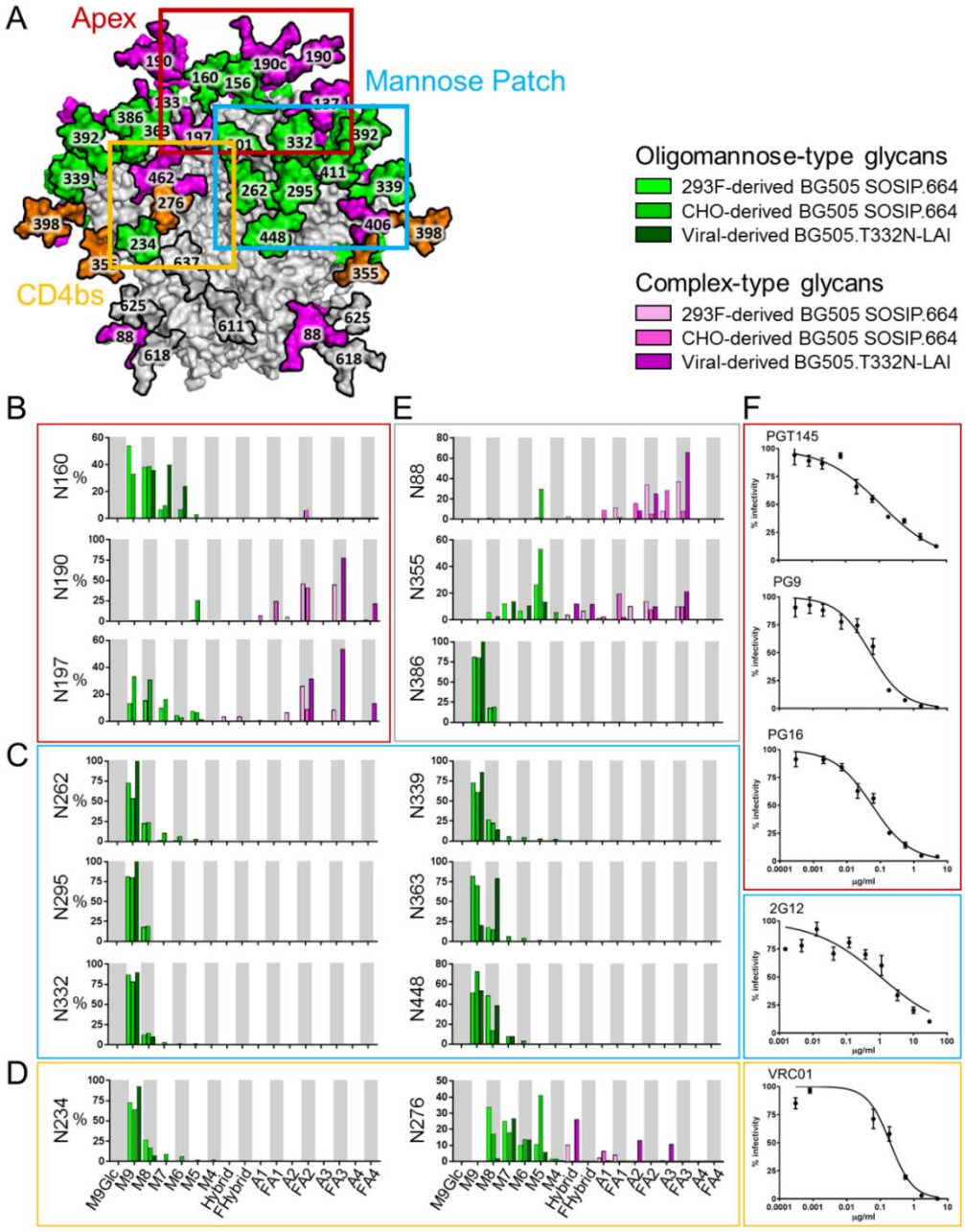


Figure 3-4: Compositional site-specific analysis of intact glycopeptides from BG505 SOSIP.664 trimers produced in 293F or CHO cells and from BG505.T332N-LAI viruses.
(A) Model of the N-glycan compositions of BG505.T332N-LAI Env highlighting three regions important for bnAb binding (reproduced from Figure 4). (B) Site-specific compositions of N-linked glycans located on the trimer apex, (C) the mannose patch, (D) proximal to the CD4 binding site (CD4bs) and (E) at remaining sites on the trimer. Data corresponding to CHO-cell derived BG505 SOSIP.664 trimers are reproduced from Dey *et al.* 2018 [109] and is included here to assist comparisons with other data sets. To simplify data presentation, complex glycans are grouped according to the presence or absence of fucose and the number of terminal antenna (e.g., FA2). (F) Neutralization by bnAbs targeting the apex (red), mannose patch (blue) and CD4bs (orange). Data are represented as mean ± SEM.

Another region of significant interest for immunogen design strategies is the trimer apex. Glycans located at the apex are highly conserved and form the epitopes for a range of bnAbs including PGT145, PG9 and PG16 [184]. The N160 glycan is present in close to 100% of HIV-1 strains and is therefore a valuable target for bnAb elicitation strategies (**Figure 3-4A**) [99]. The apex region of 293F cell-derived SOSIP.664 trimers is enriched for oligomannose-type glycans compared to gp120 monomers [79]. Using the same glycopeptide pool generated above, it was possible to investigate whether the same steric constraints upon glycan processing were present on BG505.T332N-LAI virion-derived Env. Here, the oligomannosylation seen at N160 (**Figure 3-4B**) is a strong indicator that the same steric restrictions do, in fact, apply. The high oligomannose content of the apex is also consistent with the ability of the PGT145, PG9 and PG16 bnAbs to neutralize the BG505.T332N-LAI virus (**Figure 3-4F**), as N160 mannose moieties are key components of their epitopes [106,269]. Whilst the previously reported trimer-specific restriction in glycan processing found at the conserved N160 site is conserved between the BG505.T332N-LAI-derived Env and both preparations of BG505 SOSIP.664 trimers, significant differences were observed at other sites around the apex. Thus, sites at the SOSIP.664 trimer apex that present a mixed population of oligomannose- and complex-glycans, for example N197, are more highly processed when present in the context of the BG505.T332N-LAI virus (**Figure 3-4B**). As well as an elevation in oligomannose processing, there was an increase in branching at the N197 site, and also at N190 (**Figure 3-4B**).

The CD4 binding site (CD4bs) region also contains multiple bnAb epitopes, many of which are shielded or otherwise influenced by the highly conserved N276 glycan [191]. Here, the N276 glycan was more extensively processed on the BG505.T332N-LAI virion Env than on either SOSIP.664 trimer preparation (**Figure 3-4D**). Thus, the N276 site, together with N197, N190 and N462, constitute multiple sources of glycan heterogeneity surrounding the CD4bs. The heterogeneity observed here on the virion-derived material does not generate any VRC01-resistant virions (**Figure 3-4D**), presumably because the conformational plasticity of glycans means that even large glycans do not preclude VRC01 binding.

A further feature of the BG505.T332N-LAI virus-derived material is the identification of hybrid-type glycans that are found in lesser amounts on both preparations of BG505 SOSIP.664 trimers. Specifically, greater amounts of hybrid-type glycans at N276 and N355 on the virion Env than on the SOSIP.664 trimers were observed (**Figure 3-4D** and **Figure 3-4E**). There is growing interest in hybrid-type glycans now that it is known that bnAbs such as PG16 preferentially recognize such structures [270]. The hybrid glycan populations identified here do not correspond to known bNAb epitopes, but it will be important to seek and characterize such structures in future studies of various Env proteins.

Changes at the N88 site near the base of the trimer are also noteworthy. It has been reported that the N88-glycan on gp120 from BaL virions is exclusively of the oligomannose-type [163]. This glycan is structurally proximal to the membrane, and hence it is plausible that its processing might be governed by the same factors as apply to the nearby gp41 glycans. In this study, the N88-glycan on BG505.T332N-LAI gp120 was exclusively complex, but it was significantly less processed on both SOSIP.664 trimer preparations (**Figure 3-3A, B** and **Figure 3-4E**). These observations further illustrate how proximity to the plasma membrane may influence the ability of glycan-processing enzymes to act on certain sites, irrespective of the Env clade or genotype.

3.9 Discussion

Both the global and the site-specific glycosylation of virion derived Env is largely recapitulated by recombinant SOSIP.664 trimers expressed either transiently in 293F cells or stably in CHO cells under cGMP conditions. These observations are consistent with the finding that the recombinant trimers display the epitopes for numerous glycan-targeting bnAbs that neutralize the corresponding virus and support their exploration as clinical immunogens [76,109]. However, notable points of divergence were observed between the viral and recombinant proteins that may help define how best to produce Env protein immunogens. In particular, the trimers from both 293F and CHO cells display reduced levels of glycan branching and terminal processing than their virion-derived counterparts. This trend occurs both at sites that are always fully complex-type and at mixed sites where the levels of oligomannose-type glycans are reduced on virus-derived Env. Due to the difficulties of eliciting glycan-binding bnAbs, it may be advisable to avoid exposing the immune system to glycan epitopes that are not presented on viral Env. It will be important to extend these types of glycan analysis to viruses and SOSIP trimers of additional genotypes, with the goal of refining how trimer-based immunogens are glycosylated. It will also be critical to understand the site-specific glycosylation of virus from PBMCs or primary T-cells as we can expect to discover cell-dependent influences on glycosylation patterns. The next generation of clinical immunogens may also benefit from the development of strategies to boost the immunogenicity of glycan-dependent epitopes.

Chapter 4 Enhancing glycan occupancy of soluble HIV-1 envelope trimers to mimic the native viral spike.

4.1 Contributions

The work outlined in **Chapter 4** details the combined efforts of multiple laboratories. This section includes site-specific glycan analysis but also includes the work of other collaborators as it provides context for the site-specific data and explains the rationale behind certain modifications of immunogens which would be absent if only the site-specific analysis was included. This work is a co-authored study between me and Dr Ronald Derking, with all glycan analysis performed by myself and the immunogen production and antigenic testing performed at the University of Amsterdam UMC. In addition, electron microscopy studies were used to confirm immunogen integrity, and to cross validate the observed site-specific glycan occupancy data and was performed by the laboratory of Prof. Andrew Ward (The Scripps Research Institute). Production of ExpiCHO BG505 SOSIP and surface plasmon resonance experiments were performed at in the laboratory of Prof. John Moore (Weill Cornell Medical College).

4.2 Introduction

The HIV-1 envelope glycoprotein spike (Env) initiates viral entry in host cells and is the only target for neutralizing antibodies (nAbs) [252,271]. Env is a trimer of heterodimers consisting of gp120 and gp41 subunits. During the early stages of protein synthesis Env is decorated by N-linked glycans which, upon leaving the Golgi apparatus, comprise approximately 50% of the total mass [272,273]. The N-linked glycans play a critical role in various stages of the viral life cycle, including Env folding, binding to lectin receptors, and immune evasion by shielding underlying protein epitopes [61,84]. The gp120 subunit can contain up to 35 potential N-linked glycosylation sites (PNGS) and gp41 contains typically 4 PNGS. Hence, there can be as many as 120 PNGS per trimer. Most of these PNGS are attachment points for N-linked glycans by the host cell glycosylation machinery, but some may remain unoccupied [153,155,162].

N-linked glycans are attached to asparagine residues (N) located in NxT/S sequons where T is threonine, S is serine and x can be any amino acid except proline. The amino acid that follows the NxT/S sequon can also affect glycan attachment: a proline at this position abrogates glycosylation [274]. The enzyme responsible for glycan attachment is

oligosaccharyltransferase (OST), which consists of seven to eight non-identical subunits. The catalytic subunit has two isoforms STT3A and STT3B [275,276]. The STT3A subunit is associated with the translocon and attaches glycans during translation, but it sometimes skips sequons that are in close proximity, i.e., NxT/S-NxT/S or NxT/S-x-NxT/S motifs (termed gap-0 and gap-1 sites, respectively). The STT3B subunit facilitates co-translational and post-translational glycosylation of sequons that are skipped by STT3A. Even so, not all gap-0 and gap-1 sites are fully occupied by glycans [276]. HIV-1 Env contains a number of gap-0 and gap-1 sites, such as the N156/N160 gap-1 site that contains glycans that are essential for several broadly neutralizing antibodies (bNAb) (Los Alamos database; http://www.hiv.lanl.gov) [74,269,276,277].

The protein surface of Env is almost completely shielded by glycans, which contributes to its poor immunogenicity [278]. Env and Env-based vaccines, including native-like SOSIP trimers and SOSIP-based clinical candidates, have holes in their glycan shield from two different sources. Firstly, a conserved PNGS can be absent from the Env of any particular isolate; for example, the conserved glycosylation sequons at N241 and N289 are not present in the BG505 virus and SOSIP trimers. The absence of these two glycans creates a large immunodominant hole in the glycan shield that is responsible for generating the majority of strain-specific NAb responses induced by BG505 SOSIP trimers in [219,279]. Of note is that HIV-1 isolates with an intact glycan shield are more capable of inducing broader neutralizing responses than ones with multiple holes, which may act as immunological decoys and thereby impede the development of neutralization breadth [280]. Filling glycan holes may therefore be a strategy to focus the immune response on more desirable targets; it can be achieved simply by restoring the missing PNGS [281,282]. Conversely, removing PNGS is a proven strategy to attract antibody responses to the region of choice [283–285].

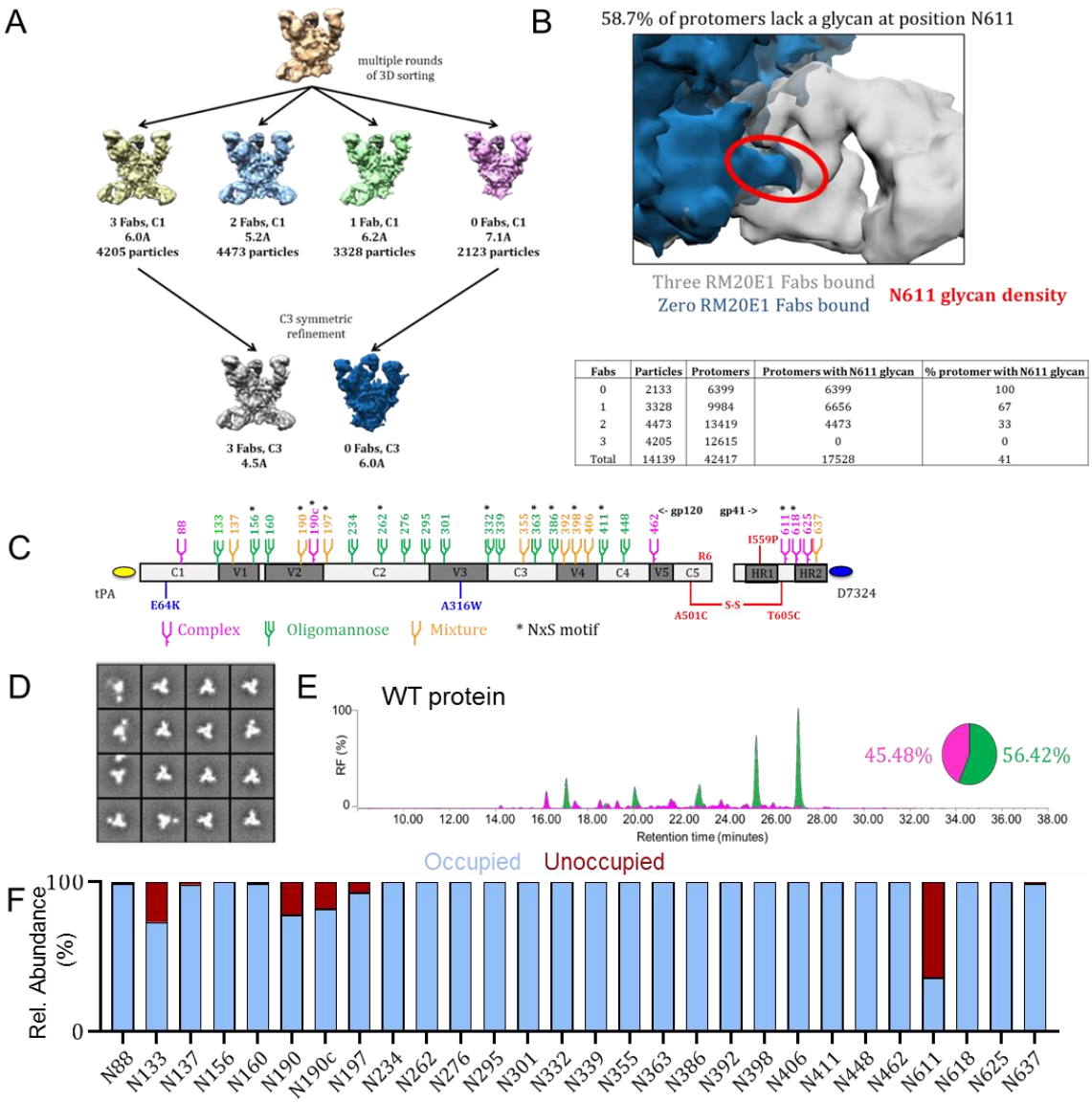
Secondly, a PNGS that is in the protein sequence can be under-occupied, likely because the site is often skipped by STT3A and STT3B [275,276,286]. Several PNGS in the V1/V2 domain of BG505 SOSIP.664 trimers are under-occupied, creating artificial glycan holes [153]. The PNGS at position 611 of these trimers is probably also under-occupied, because it induces antibodies that can only neutralize viruses from which N611 is deleted [218,219,279]. Direct comparisons revealed that glycan occupancy is lower on BG505 SOSIP trimers than on the sequence-matched viral Env [153,264]. However, increasing the occupancy of a PNGS that is likely caused by OST skipping is not straightforward as the factors that determine whether a PNGS becomes a substrate for OST are poorly understood. Studies on model glycoproteins have, however, shown that NxT sequons have a two- to three-fold higher probability of becoming glycosylated compared to NxS [287], most likely because OST has a higher affinity for NxT [288,289]. This study aims to increase PNGS occupancy on prototypic BG505 SOSIP trimers to an extent that mimicked viral BG505 Env, and thereby

eliminate artificial glycan holes. This should facilitate immuno-focusing to more relevant epitopes.

4.3 Cryo-EM confirms N611 under-occupancy

Potential N-linked glycan site occupancy analyses of soluble trimers and their viral counterparts are less common, but new techniques have enabled such studies on BG505 SOSIP.664 and other trimers [153,162,264,290]. Here, cryo-electron microscopy (cryo-EM) was performed as well as PNGS occupancy and glycan composition analyses on BG505 SOSIP.v4.1 trimer (from here on referred to as WT proteins). These trimers have additional mutations to aid with the stability of the trimer, and are more commonly used in recent studies involving BG505 SOSIP [291]. Monoclonal antibodies (mAbs) have been isolated from rabbits and rhesus macaques that were immunized with BG505 SOSIP.664 trimers by single B-cell sorting and B-cell receptor (BCR) cloning [218,219,279]. While many of the mAbs targeted the N241/N289 glycan hole, a subset of mAbs were directed against the N611 artificial glycan hole. These mAbs could bind the WT proteins, but were unable to neutralize the corresponding virus unless the N611 PNGS was knocked-out [153,218,219,279]. The implications of these findings are that the region underneath the N611 glycan on BG505 SOSIP trimers is immunogenic for mAb induction because the N611 PNGS is under-occupied; and that the same site is more extensively occupied on the mAb-resistant virus.

To probe the N611 under-occupancy on the BG505 SOSIP trimer, mAb RM20E1, directed to the N611 site, which was isolated from a BG505 SOSIP.664-immunized macaques, was used to perform cryo-EM studies with the Fab in complex with the BG505 SOSIP.v5.2 trimer [92,218,219]. Multiple rounds of 3D sorting were performed to segregate trimers with zero, one, two, or three RM20E1 Fabs attached (**Figure 4-1A**). Density corresponding to the N611 glycan was observed on all protomers with no Fab attached, but in contrast there was no visible N611 density on any protomers bound to RM20E1 (**Figure 4-1B**). The Fab also makes contacts with the fusion peptide and adjacent protomers [218]. These observations are consistent with the previously described binding and neutralization results [218,279]. A further analysis of the numbers of trimers in 3D-sorted classes with zero, one, two, or three RM20E1 Fabs showed that N611 was occupied on ~40% of the protomers.



Fabs	Particles	Protomers	Protomers with N611 glycan	% protomer with N611 glycan
0	2133	6399	6399	100
1	3328	9984	6656	67
2	4473	13419	4473	33
3	4205	12615	0	0
Total	14139	42417	17528	41

Figure 4-1: Several PNGS on BG505 SOSIP.v4.1 trimers are under-occupied.
(A) Cryo-EM analysis of WT proteins in complex with the RM20E1 Fab. Complexes were sorted based on the number of Fabs bound; the numbers of particles in each reconstruction are listed, as are the resolutions of the final reconstructions. (B) Overlay of the Cryo-EM reconstructions of the BG505 SOSIP.664 trimer alone (in blue) and in complex with the RM20E1 Fab (in grey). The density of the N611 glycan (also in blue) on the trimer without RM20E1 is highlighted in red to illustrate clashes with the RM20E1 Fab (in grey). The total number of complexes with different numbers of Fab bound are indicated in the table, as are the protomers with N611 glycan density. From these numbers, the % of protomers that lack the N611 glycan could be calculated (~59%). Electron microscopy was performed by Prof. Ward laboratory (Scripps Research Institute) (C) Linear representation of the D7324-tagged BG505 SOSIP.v4.1 trimer. The SOSIP changes and the stabilizing E64K and A316W mutations are all highlighted [76,291]. The glycans are also indicated, with the amino acid numbering based on the HXB2 strain. Sites that are predominantly occupied by oligomannose species are coloured green and by complex species in magenta, while orange indicates that either an oligomannose or a complex glycoform can be present, based on data presented in Figure 4-1D. An asterisk indicates that the PNGS is encoded by an NxS sequon. (D) NS-EM analysis of WT proteins, showing the 2D class-averages. (E) HILIC-UPLC analysis of WT proteins. Peaks coloured green represent glycans that can be cleaved by endoglycosidase H (endoH) and correspond to oligomannose or hybrid-type glycans. The HILIC-UPLC spectra and pie chart show the overall oligomannose glycan (green) and complex/hybrid glycan (magenta) content. (F) Quantification of site-specific

occupancy for all 28 PNGS on WT proteins, derived from LC-ESI MS analyses. Results are the mean of two independent biological replicates of the WT protein. The data displayed represents the relative proportion of occupied (blue) vs unoccupied (red). All the data in panels A-F were derived using WT proteins produced in HEK293F cells followed by PGT145-affinity purification. The corresponding data on WT proteins purified by the 2G12/SEC method are in Appendix E.1.

4.4 Several PNGS on BG505 SOSIP trimers are under-occupied

Next, several techniques were used to gain a more detailed understanding of the glycan shield on the WT protein. The D7324-tagged WT protein (**Figure 4-1C**) was expressed in HEK293F cells and purified by affinity chromatography using the PGT145 bNAb, which is selective for native-like trimers [292]. In addition, affinity purified D7324-tagged WT protein was purified using the 2G12 bNAb followed by size exclusion chromatography (SEC) [76,291]. Both methods are routinely used for purifying native-like trimers, including in cGMP programs [109,292]. WT proteins purified via the PGT145 and 2G12/SEC methods had a homogeneous native-like conformation as judged by negative stain electron microscopy (NS-EM) (**Figure 4-1D** and **Appendix E.1**). Overall, analyses of the PGT145- and 2G12/SEC purified-proteins yielded very similar data, with a few notable exceptions (see below). Data on the PGT145-purified proteins is presented in the main Figures, with the corresponding Figures for their 2G12/SEC-purified counterparts in the **Appendix**.

To determine the total oligomannose content of the purified WT protein, hydrophilic interaction liquid chromatography-ultra performance liquid chromatography was performed (HILIC-UPLC). The PGT145-purified WT protein contained 56% oligomannose-type glycans and 60% when purified using 2G12/SEC. In both cases, $Man_8GlcNAc_2$ and $Man_9GlcNAc_2$ were the predominant glycan species (**Figure 4-1E** and **Appendix E.1B**), consistent with previous reports on native-like SOSIP trimers [100,153,291].

To investigate PNGS occupancy and glycan composition, site-specific analysis of the WT protein using liquid chromatography-electrospray ionization (LC-ESI) mass spectrometry (MS) on an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) was used (**Figure 4-1F**). PNGS occupancy is expressed as the percentage of the total peptide that is modified by a glycan; with a modification level of ≥ 95% as indicative of full occupancy. The majority of PNGS on gp120 from the PGT145-purified WT protein were fully occupied, including the five located within a 25-amino acid stretch of V4 (N386, N392, N398, N406, and N411). However, under-occupancy was seen for sites near the apex: N133: ~75% occupancy; N190: ~80%; N190c: ~80%; and N197: ~93% (**Figure 4-1F**). The sites under-occupied on the 2G12/SEC-purified WT protein were: N133: ~90% occupancy, N137: ~80%, N190: ~90%; and N190c: ~80% (**Appendix E.1C**). These regions of under-occupancy are similar to those described previously for BG505 SOSIP.664 trimers [153].

The N611 glycan on the PGT145-purified protein is markedly under-occupied (~35% occupancy) (**Figure 4-1F** and **Appendix E.1C**). This finding is highly concordant with the cryo-EM analysis (**Figure 4-1A** and **B**). The other three gp41 sites (N618, N625 and N637) were fully occupied on the PGT145-purified WT protein. However, the N625 site was less occupied (~70%) and the N611 site more occupied (70%) on the 2G12/SEC-purified WT protein (**Appendix E.1C**). The LC-ESI MS method also enabled the determination of the specific glycoforms present at each site. For most PNGS on the WT protein, the glycoforms were very similar to those observed in previous studies [100,153] (**Appendix E.1D**). Thus, sites N156, N160, N234, N262, N276, N295, N332, N339, N363, N386, N392 N411 and N448 were predominantly occupied by oligomannose glycans, sites N190c, N398, N611, N618 and N625 by complex glycans, and sites N88, N137, N190, N197, N301, N355, N406, N462 and N637 by a mixture of both forms.

4.5 Glycan occupancy is enhanced by PNGS sequon engineering

Although the determinants of glycan occupancy are not clear, studies on model glycoproteins have shown that the NxT sequon has a higher probability of becoming glycosylated compared to an NxS sequon [287]. The use of NxS (~40%) versus NxT (~60%) in the BG505 sequence is consistent with most isolates in the Los Alamos database (**Table 1**). Thus, when an NxS sequon is present at a particular site, it is also the case for BG505. To increase glycan occupancy, two stabilized BG505 SOSIP.v4.1 trimers were designed in which all the PNGS were changed to either NxT or NxS. The two variants are designated NxT protein and NxS protein, respectively.

Table 1: Frequency of NxS/T in natural isolates

			Frequency in natural isolates [1]			
Glycan	**BG505**	**Motif**	**% NxT**	**% NxS**	**% No PNGS**	**Glycan occupancy**
N88	NxT	NVT	98.2	0.4	1.4	>95%
N133	NxT	NVT	15.0	3.0	82.0	75%
N137	NxT	NIT	20.1	3.5	76.4	>95%
N156	**NxS**	**NCS**	**8.6**	**87.6**	**3.8**	**>95%**
N160	NxT	NMT	88.9	2.1	9.0	>95%
N190	**NxS**	**NRS**	**N.D [2]**	**N.D [2]**	**N.D [2]**	**80%**
N190c	**NxS**	**NNS**	**N.D [2]**	**N.D [2]**	**N.D [2]**	**75%**
N197	**NxS**	**NTS**	**1.1**	**96.8**	**2.1**	**90%**
N234	NxT	NGT	74.8	5.0	20.2	>95%
N262	**NxS**	**NGS**	**0.2**	**98.9**	**0.9**	**>95%**
N276	NxT	NIT	76.5	18.5	5.1	>95%
N295	NxT	NCT	58.2	1.5	40.4	>95%
N301	NxT	NNT	93.7	0.2	6.2	>95%
N332	**NxS**	**NVS**	**3.0**	**68.5**	**28.5**	**>95%**
N339	NxT	NET	65.1	0.1	34.8	>95%
N355	NxT	NNT	72.8	1.6	25.5	>95%
N363	**NxS**	**NSS**	**0.4**	**7.7**	**91.9**	**>95%**
N386	**NxS**	**NTS**	**40.0**	**46.8**	**13.2**	**>95%**
N392	NxT	NST	70.0	9.2	20.8	>95%
N398	**NxS**	**NTS**	**11.0**	**2.9**	**86.1**	**>95%**
N406	NxT	NST	2.7	0.5	96.8	>95%
N411	**NxS**	**NDS**	**44.4**	**1.3**	**54.3**	**>95%**
N448	NxT	NIT	87.4	0.1	12.5	>95%
N463	NxT	NST	N.D [2]	N.D [2]	N.D [2]	>95%
N611	**NxS**	**NSS**	**11.6**	**86.8**	**1.6**	**40%**
N618	**NxS**	**NLS**	**16.6**	**75.3**	**8.1**	**>95%**
N625	NxT	NMT	96.2	0.0	3.8	>95%
N637	NxT	NYT	92.5	3.8	3.7	>95%

[1] Data obtained from 6516 unique HIV envelope sequences found in the Los Alamos database (http://www.hiv.lanl.gov)
[2] PNGS are located in highly variable regions and the specific sites differ between the natural isolates, making reliable calculations not possible.
Bold indicates the PNGS that have an NxS motif

SDS-PAGE and Blue Native (BN-) PAGE was used to assess the expression and trimer formation of unpurified WT, NxT and NxS proteins produced in transfected HEK293T cells (**Appendix E.1D** and **E**). The NxT protein was expressed efficiently and formed trimers, although slightly less well than the WT protein, but the NxS protein was not expressed efficiently. A D7324-capture ELISA using bNAbs 2G12, VRC01, PGT145 and PGT151 to probe the folding of the proteins in the unpurified supernatants confirmed these findings (**Appendix E.1F**). Thus, all four bNAbs bound to the NxT and WT proteins, but the NxS protein could only be detected using 2G12. The defect in NxS protein expression might be caused by the absence of glycans important for Env to fold properly in the ER [293].

Purified, HEK293F cell expressed NxT proteins were fully cleaved with no evidence of aggregation or dissociation (**Appendix E.1**). Compared to WT, the NxT protein migrated more slowly through the SDS- and BN-PAGE gels, indicating more extensive glycosylation. The NS-EM 2D-class averages showed that the NxT protein was in a native-like conformation (**Figure 4-2A** and **Appendix E.2A**).

The overall glycan composition of the NxT protein was similar to WT, with oligomannose contents of 56% and 60% after PGT145 and 2G12/SEC-purification, respectively (**Figure 4-2B** and **Appendix E.2B**). However, the LC-ESI MS experiments revealed some striking differences in PNGS occupancy. Whereas the N190, N190c and N197 PNGS were under-occupied on the WT protein (**Figure 4-1F**), the same sites were >95% occupied on both preparations of the NxT proteins (**Figure 4-2C** and **Appendix E.2C**). Furthermore, occupancy of N611 on gp41 increased from ~40% and ~70% on the PGT145- and 2G12/SEC-purified WT proteins, respectively, to ~90% and ~85% for the corresponding NxT proteins (**Figure 4-2C** and **Appendix E.2C**). Occupancy of some sites including N392, N398, N406 and N625 could not be determined. These sites tend to be in regions that are densely glycosylated and ionize poorly, resulting in glycopeptides of insufficient quality. Other studies encountered similar problems previously [153,264].

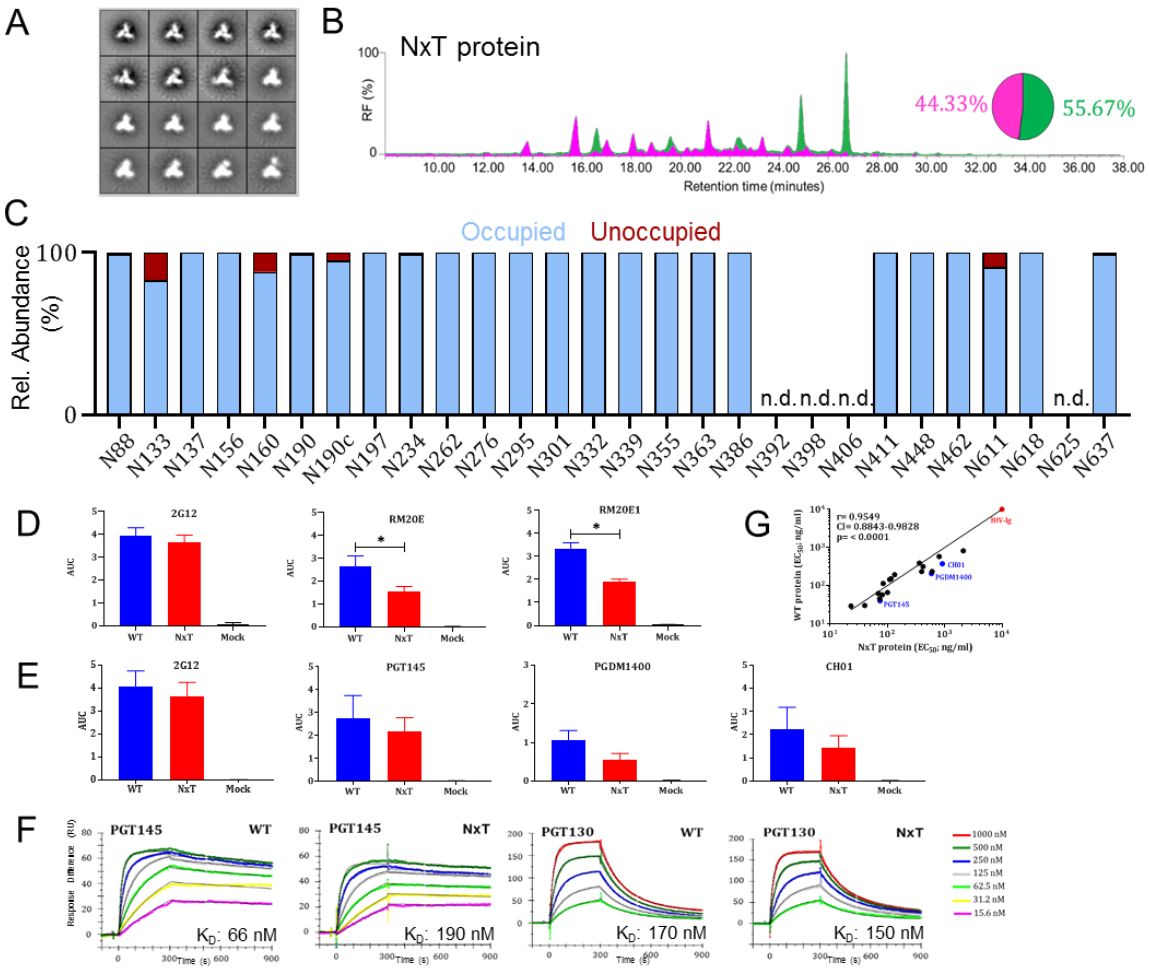


Figure 4-2: Glycan occupancy is increased by PNGS sequon engineering.
(A) NS-EM analysis of NxT proteins, showing the 2D class-averages, performed by the laboratory of Prof. Ward (The Scripps Research Institute). (B) HILIC-UPLC analysis of the NxT protein. The colour coding of the spectra and pie chart is the same as in Figure 1E. (C) Quantification of site-specific occupancy for all 28 PNGS on NxT proteins, derived from LC-ESI MS analyses. (D) Binding of WT and NxT proteins to two N611-directed non-NAbs, RM20E and RM20E1, isolated from BG505 SOSIP.664 immunized rhesus macaques. The area under the curve (AUC) values derived from ELISA titration curves are plotted. * indicates a significant difference ($P < 0.05$) between the WT and NxT proteins, calculated using a Mann-Whitney 2-tailed test. (E) Binding of WT and NxT proteins to three V2-apex

directed bNAbs, PGT145, PGDM1400 and CH01, and 2G12 for comparison. AUC values derived from derived from ELISA titration curves are plotted. (F) Binding of bNAbs PGT145 and PGT130 to WT and NxT proteins, assessed by SPR. (G) The EC50 values derived using WT and NxT proteins were plotted and compared using Spearman's correlation coefficient. All analyses were performed on NxT proteins produced in HEK293F followed by PGT145 purification. The corresponding data on NxT proteins purified by the 2G12/SEC method are in **Appendix E.2**. Antigenic characterization displayed in panels D, E, F and G were performed in the laboratories of Prof. Rogier Sanders and Prof. John Moore.

Occupancy at N133 was not restored on the PGT145-purified NxT protein as it increased from 73% to 83%, which means that ~1 in 5 protomers still lack a glycan at this position. Conversely, N133 occupancy decreased to ~60% on the 2G12/SEC-purified NxT protein, compared to ~95% for WT (**Appendix E.2C**). Furthermore, while the N160 PNGS was fully occupied on the PGT145-purified WT protein, occupancy decreased to ~85% for the NxT protein (**Figure 4-2C**). An even greater decrease of 30 percentage points was observed for the 2G12/SEC-purified NxT protein (**Appendix E.2C**). The PGT145 epitope requires at least two occupied N160 sites, which probably explains the higher N160 occupancy on the PGT145- versus 2G12/SEC-purified NxT protein [269].

There were no substantial differences when the glycoforms at each PNGS of the NxT protein were compared to the WT protein. The sites that were more occupied on the NxT protein than on WT (N190, N190c, N197 and N611) were all predominantly occupied by complex glycans (**Appendix E.2C**). The increased occupancy of N611 on NxT proteins could reduce the immunogenicity of non-NAb epitopes associated with this artificial glycan hole. To assess this *in vitro*, the binding of two N611-directed non-NAbs RM20E and RM20E1 was determined. Using a D7324-capture ELISA, the PGT145-purified NxT protein was significantly less reactive than the WT protein with non-NAbs RM20E (P=0.03) and RM20E1 (P=0.03), calculated using a two-tailed Mann-Whitney test (**Figure 4-2D**).

The antigenicity of the PGT145-purified NxT proteins was assessed by ELISA (**Appendix E.3**). Most of the bNAbs tested recognize epitopes involving glycans; including ones directed against quaternary-structure dependent V1/V2-apex sites (PG9, PG16, PGT145, PGDM1400 and CH01), the N332-glycan dependent epitope cluster (PGT121-123, PGT125-128, PGT130, 2G12 and PGT135), the quaternary-structure dependent epitopes at the gp120-g41 interface (PGT151 and 35O22), and gp41 (3BC315), the CD4bs (VRC01 and 3BNC60) and polyclonal HIV-Ig. In all cases, the bNAbs bound similarly to the WT and NxT proteins, with <3-fold difference in 50% binding concentrations (EC50) (**Appendix E.3**). An area under the curve (AUC) analysis, however, revealed that bNAbs against the trimer apex were less reactive with the NxT protein (**Figure 4-2E**). This finding is concordant with the reduced occupancy of N160 on this trimer, as the epitopes involve this glycan. Similar results were obtained when the 2G12/SEC-purified WT and NxT proteins was tested by ELISA (**Appendix E.2D**), and by surface plasmon resonance (SPR)

(WT, KD 66 nM versus NxT, KD 190 nM for PGT145; the KD values are for the initial interaction as fitted with a conformational-change model [294] (**Figure 4-2F** and **Appendix E.2E**). The EC50 values for bNAb binding to the WT and NxT proteins were highly correlated (Spearman r=0.95, P<0.0001 (**Figure 4-2G**) and Spearman r=0.97, P<0.0001 (**Appendix E.2F**)).

To verify whether NxT sequon engineering was compatible with Env function, a BG505.T332N infectious molecular clone (IMC) was generated with all Env PNGS changed to NxT. The NxT and WT viruses were equally infectious (**Appendix E.4A**). In contrast, an IMC with the PNGS all changed to NxS was not infectious, which is consistent with the poor expression of the corresponding NxS protein (**Appendix E.1** and **E.4A**). In a neutralization assay, the NxT virus had reduced sensitivity to PGT145 (26-fold increase in IC50; Appendix **E.4C** and **D)**. A second V1/V2-apex bNAb, CH01, neutralized the NxT and WT viruses comparably, as judged by IC50 values, but the maximum extent of neutralization was only ~60% for the NxT virus while its WT counterpart was completely neutralized at high CH01 concentrations. (**Appendix E.4C**). The impact of the NxT modification on the V1/V2-apex are therefore seen with viruses and recombinant trimers and are consistent with under-occupancy of the PNGS at N160 in both contexts.

4.6 Occupancy of gap-1 sites can be enhanced by reducing OST affinity of the first site

The reduced N160 occupancy on the NxT protein was surprising since the glycan sequon was unchanged compared to WT. However, the preceding N156 sequon was changed from NxS to NxT. This site is fully occupied on WT proteins purified by both methods (**Figure 4-1F** and **Appendix E.1C**). This under occupancy may appear due to enhancing the affinity of the PNGS at N156 for OST by changing it to NxT, and subsequently interfere with glycan attachment at N160. We therefore reverted T158 to S on the NxT protein. The purified NxT T158S protein adopted a native-like conformation (**Figure 4-3A** and **Appendix E.5A**) and had an overall oligomannose content of 50% and 48% for PGT145- and 2G12/SEC-purified proteins, respectively (**Figure 4-3B** and **Appendix E.5**). The T158S mutation enhanced N160 occupancy to >95% for both preparations of purified proteins (**Figure 4-3C** and **Appendix E.5C**), similar to N160 on the WT protein (**Figure 4-1D** and **Appendix E.1C**).

The only site that remained considerably under-occupied on the NxT T158S protein was N133 (~75% occupancy; **Figure 4-3C**). Based on success with the N160 site and noting that the N133 and N137 sequons are spaced similarly to N156 and N160, the same strategy was employed to increase N133 occupancy. Accordingly, residue 135 was changed from T to S in the background of the NxT T158S protein. Occupancy of the N133

site increased from ~75% for the NxT T158S protein (**Figure 4-3C**) to ~95% on NxT T135S T158S (**Figure 4-3D, E** and **Appendix E.6A, B**). Every PNGS was now >95% occupied on the NxT T135S T158S protein (**Figure 4-3D** and **Figure 4-4**). However, for 2G12/SEC-purified NxT T135S T158S, N133 and N190 occupancy remained lower (~65% and 86%, respectively; **Appendix E.6B**). Hence, the affinity purification method can impact PNGS occupancy, but the glycan composition remains similar using different purification methods.

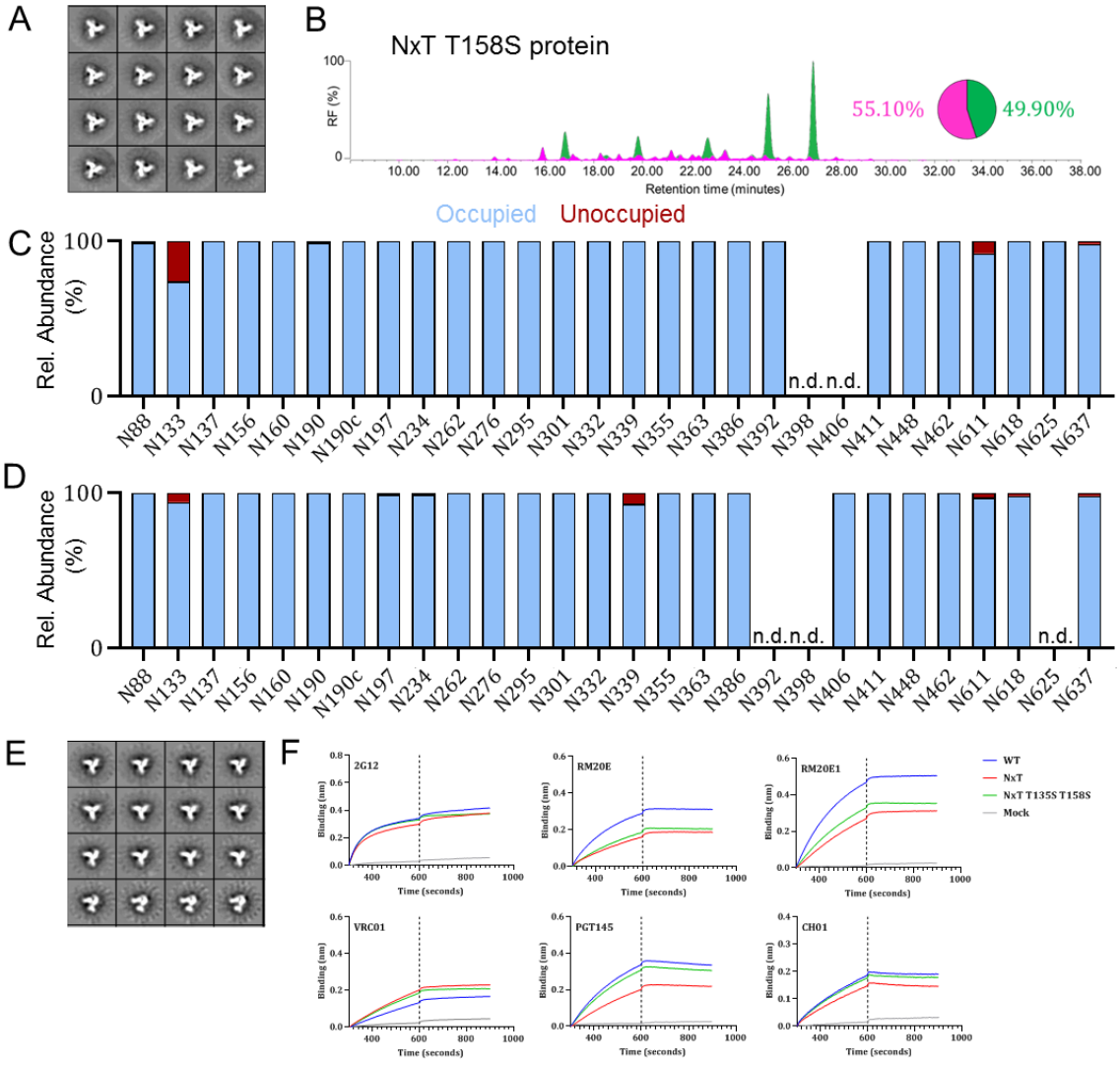


Figure 4-3: Occupancy of gap-1 sites can be increased by reducing the affinity of the first site for OST.
(A) NS-EM analysis of NxT T158S proteins, showing the 2D class-averages. (B) HILIC-UPLC analysis of the NxT T158S protein. The colour coding of the spectra and pie chart is the same as in **Figure 4-1B**. (C) Quantification of glycan occupancy using LC-ESI MS of the 28 PNGS on NxT T158S proteins. (D) Quantification of site-specific occupancy for all 28 PNGS on NxT T135S T158S proteins, derived from LC-ESI MS analyses. Results shown are the mean of two independent biological replicates of the NxT T135S T158S protein. (E) NS-EM analysis of NxT T135S T158S proteins, showing the 2D class-averages. (F) Binding of non-NAbs RM20E and RM20E1 to WT, NxT and NxT T135S T158S proteins, assessed by BLI. The bNAb 2G12 was also tested, for comparison. Additionally, the binding of bNAbs VRC01, PGT145 and CH01 was tested. The average binding curves from 3 independent duplicates are shown. All analyses were performed on NxT T135S T158S trimers produced in HEK293F cells and affinity purified using PGT145. Electron microscopy was performed

by the laboratory of Prof. Ward and binding studies by the laboratory of Prof. Rogier Sanders.

The glycoforms at each site on the PGT145-purified NxT T135S T158S and WT proteins were very similar, but with a few exceptions. The N88, N190 and N406 glycans were more processed and closer resemble what is found on viral Env (**Appendix E.6B**) [153]. Conversely, the N133 glycan was less processed on this protein than WT. The glycans at N137, N276 and N406 were also more processed on the 2G12/SEC-purified NxT T135S T158S protein (**Appendix E.6B**). The increased processing at N276 is similar to what was seen with the similarly purified NxT T158S protein but differs from their PGT145-purified counterparts.

In a Biolayer interferometry (BLI) assay, mAbs RM20E and RM20E1 were less reactive with the PGT145-purified NxT T135S T158S protein than WT, which is further evidence that N611 is more occupied on the modified trimer (**Figure 4-3F**). BLI also showed that the apex-targeting bNAbs PGT145 and CH01 were equally reactive with the WT and NxT T135S T158S proteins (**Figure 4-3F**). Thus, increasing N160 occupancy via the T158S mutation improved the presentation of bNAb epitopes at the trimer apex. In summary, the PGT145-purified BG505 NxT T135S T158S trimer closely mimics viral BG505 Env in respect of overall PNGS occupancy [153]. In particular, the artificial glycan hole recognized by non-NAbs and caused by under-occupancy of the N611 site is now closed, which should facilitate immune focusing to more relevant neutralizing epitopes.

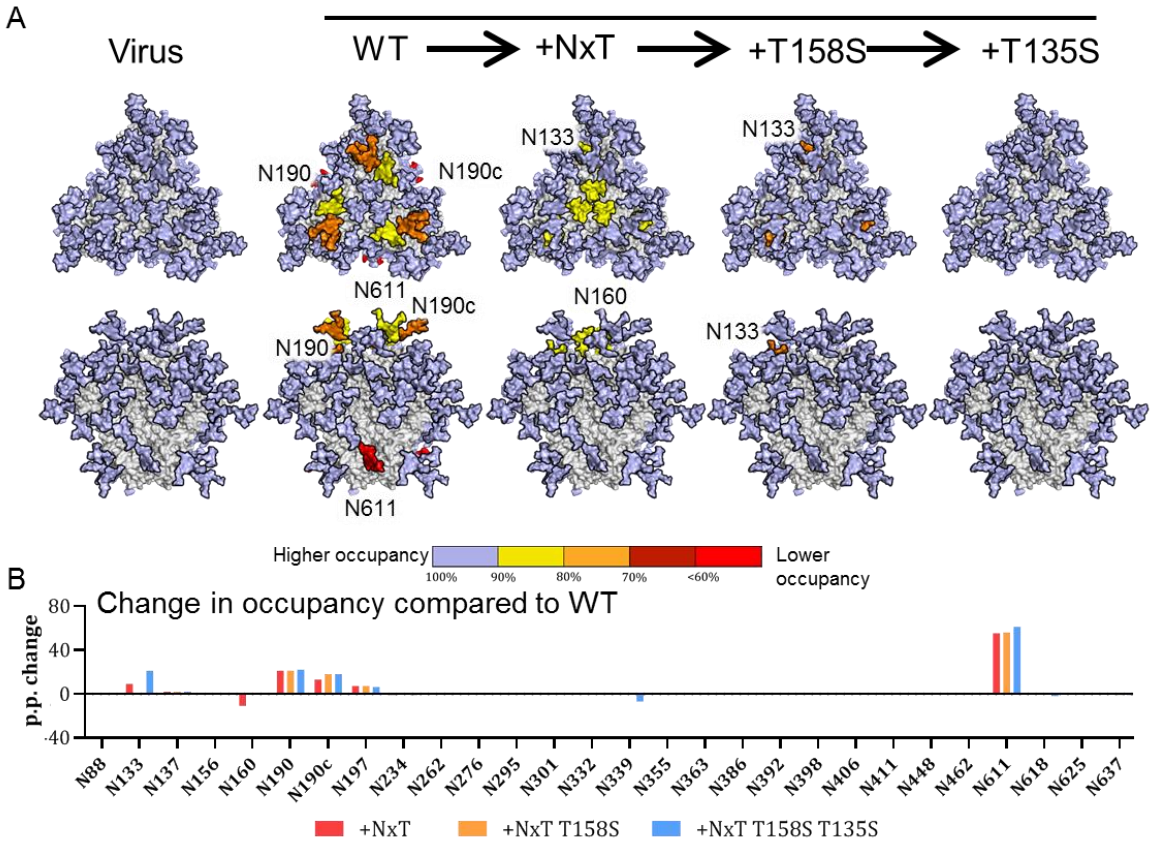


Figure 4-4 Sequential improvement of PNGS occupancy on SOSIP trimers.
(A) Occupancy of each PNGS on the sequentially improved SOSIP.v4.1 trimers (WT, NxT, NxT T158S, NxT T135S T158S) is compared with the same site on the BG505 virus. Glycans are modelled onto each PNGS on a 3D model of the WT protein. Colour coding reflects the occupancy of each PNGS: light blue, full occupancy (>95%); yellow, 80-95% occupancy; orange, 70-80% occupancy, dark red, 60-70% occupancy; bright red, <60% occupancy. (B) The sequential changes in PNGS occupancy at each PNGS are shown by the bars. The data shown are the percentage changes in occupancy caused by the various PNGS sequon changes, compared to the same site on the WT BG505 SOSIP.664 trimer.

4.7 PNGS sequon engineering translates to cells from different species

While the work described above was performed using human HEK293F cells, these cells are not commonly used for clinical grade proteins, although they have been used to produce recombinant human clotting factor VIII for clinical trial usage [109,295,296]. To investigate the effect of PNGS sequon engineering in cells more relevant for production of clinical grade material, ExpiCHO-S cells were transfected with the WT protein and affinity-purified it is using PGT145. LC-ESI MS analyses revealed that the same sites that were underoccupied in HEK293F WT protein were also under-occupied on the ExpiCHO-S WT protein: N133, ~85% occupancy; N190c, ~62%; N197, ~66%; and N611 site on gp41 (~37% occupancy; **F**igure **4-5A** and **Appendix E.6C, D**). In addition, a number of sites were less occupied on the ExpiCHO-S material compared to the HEK293F PGT145-purified WT protein: N137, ~85%; and N156, ~92%; and most notably N190, ~8%; and N625, 0%

occupancy (**Appendix E.6C, D**). N137 and N625 were also under-occupied on HEK293F WT produced proteins purified using 2G12/SEC-purification, but not to the extent observed with ExpiCHO-S. Glycoforms at each site were consistent between the cell lines with a few exceptions. Increased processing was observed at N197 and N406 whereas N276 and N637 possessed a higher abundance of oligomannose-type glycans on the ExpiCHO-S material.

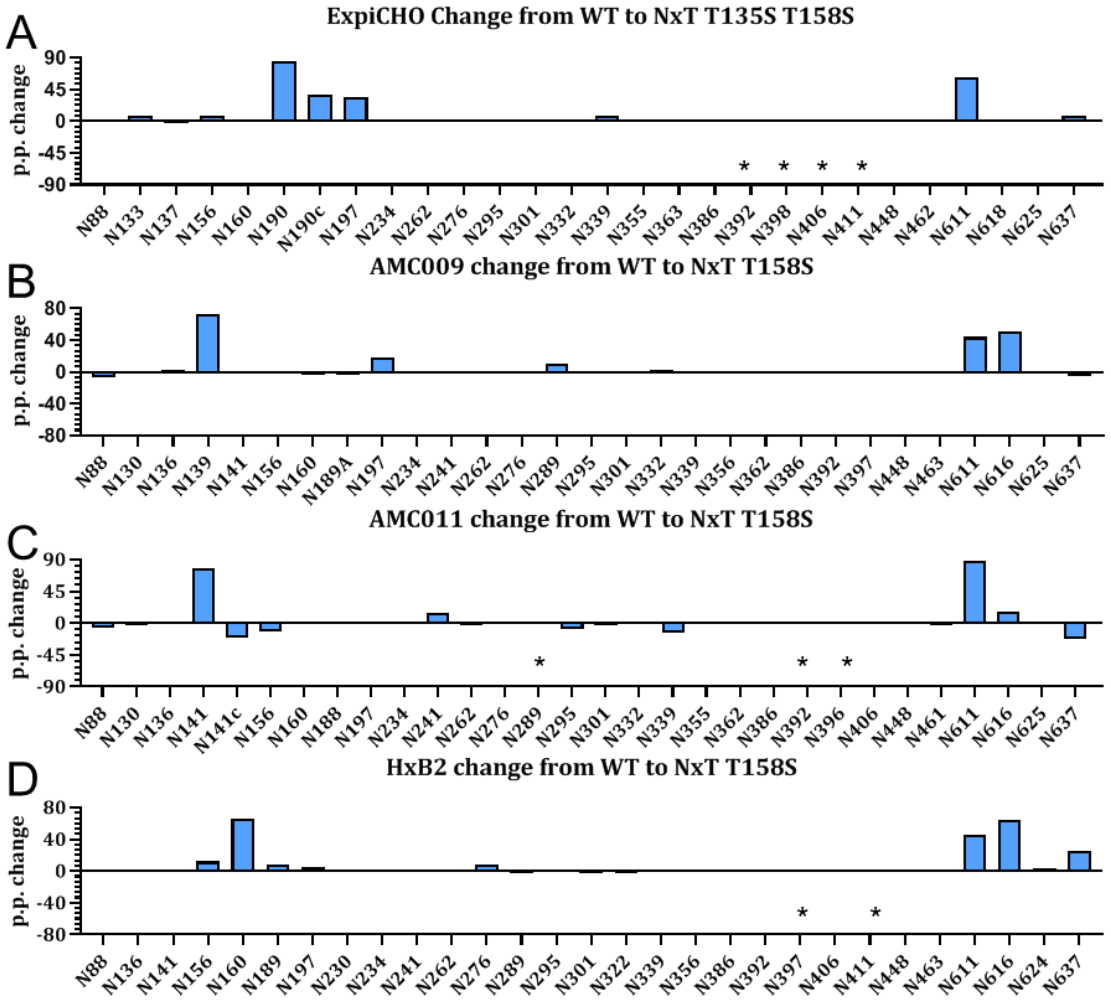


Figure 4-5: PNGS sequon engineering is applicable to Env trimers produced in cells from different species and derived from diverse HIV-1 strains.
(A) Change in PNGS occupancy between WT proteins produced in ExpiCHO-S cells upon introduction of NxT T135 T158S. The presented data represents the arithmetic difference between the glycan occupancy of the NxT protein minus the WT glycan occupancy, representing a percentage point change (p.p.). A positive p.p. change represents a higher occupancy of the NxT variant compared to the WT. Sites labelled with an asterisk represent those for which data could not be obtained for either the WT or the NxT variant. For additional Env sequences the data is displayed in an identical manner for (B) AMC009, (C) AMC011 and (D) HxB2. N411 was introduced into HxB2 NxT but is not present in WT and so is marked as an asterisk.

PNGS occupancy of ExpiCHO-S produced protein was higher on the NxT T135S T158S protein compared to the corresponding WT protein. Occupancy of PNGS N156, N190c and N197 was restored fully (>95% occupancy), whereas occupancy of N133 and N190 increased to ~90% (**F**igure **4-5A** and **Appendix E.6C, D**). The N137 site remained unchanged and under-occupied (~85% occupancy). On gp41, the occupancy of N611 was restored (>95% occupancy), but N625 remained completely under-occupied (0% occupancy). Overall, the benefit of PNGS sequon engineering for HEK293F-expressed protein was reproduced in ExpiCHO-S cells. However, an important observation is that PNGS under-occupancy was more pronounced on ExpiCHO-S produced protein. Thus, in particular the two sites at N190 and N625 were notably less occupied on ExpiCHO-S material. The under-occupancy of N625, which is present as an NxT sequon in WT and mutant proteins, was not restored in the NxT T135S T158S protein. N625 occupancy can vary between protein preparations and other approaches should be taken to address this.

4.8 PNGS sequon engineering enhances PNGS occupancy on SOSIP trimers from diverse isolates

To explore whether PNGS sequon engineering could be applied to Env trimers from other virus isolates, three additional SOSIP trimers containing NxT T158S mutations were generated. Two were based on clade B sequences AMC009 and AMC011 [297,298] and the third was based on the HxB2 reference sequence. These trimers were transiently expressed, for both WT and NxT T158S proteins, in HEK293F cells, followed by PGT145-affinity purification for AMC009 and AMC011 and 2G12/SEC-purification for HxB2 (**Appendix E.7**). The antigenicity of these trimers was assessed using bNAbs 2G12, VRC01, PGT145 and PGT151 (**Appendix E.7B** and **C**). Overall, the bNAbs bound similarly to the WT and NxT T158S proteins. PGT145 did not bind the HxB2 and HxB2 NxT T158S proteins.

LC-ESI MS analyses revealed that on the AMC009 WT protein the following sites were under-occupied: N139, ~15% occupied; N156, ~90%; N197, ~75%; N289, ~10%; N611, ~40%; N616, ~45%; and N637, ~35% (**F**igure **4-5B** and **Appendix E.7D**). PNGS occupancy on the AMC011 WT protein overall was higher at each site, compared to the other proteins, but a few sites were highly under-occupied: N141, ~20% occupied; N241, ~85%; N611, ~10%; and N616, ~5% (**F**igure **4-5C** and **Appendix E.7D**). For the HxB2 WT protein the following sites were under-occupied: N156, ~55% occupied; N160, ~25%; N189, ~45%; N262, ~90%; N611, ~45%; N616, ~15%; and N637, ~60% (**F**igure **4-5D** and **Appendix E.7D**). In general, the PNGS that are most frequently underoccupied are those surrounding the trimer apex and those on gp41. This is very similar as observed for the BG505 WT protein, but with isolate-specific differences (**Figure 4-1F**).

PNGS occupancy was enhanced substantially on the AMC009, AMC011 and HxB2 NxT T158S proteins (**F**igure **4-5** and **Appendix E.7D**). On the AMC009 NxT T158S protein, occupancy of all PNGS approached full occupancy, except N637 in gp41 for which occupancy did not change (**F**igure **4-5B** and **Appendix E.7D**). On the AMC011 NxT T158S protein, the most notable changes were seen for N141, N241 and N611 that reached full occupancy (>95% occupancy) (**F**igure **4-5C** and **Appendix E.7D**). Occupancy of N616 on the AMC011 NxT T158S protein increased, from ~10% to ~20% occupied, and PNGS occupancy at N637 decreased, from ~85% to ~60% occupied. On the HxB2 WT protein, the N156 and N160 sequons are both NxS and on the HxB2 NxT T158S protein the occupancy of N156 and N160 increased (to ~70% and ~89% occupancy, respectively) (**F**igure **4-5D** and **Appendix E.7D**). This confirms that the optimal sequon combination at the apex is NxS-x-NxT. Occupancy of N189 was not restored, but occupancy of the N611 and N616 sites on gp41 approached full occupancy. The N637 sequon was also unchanged on the NxT T158S protein, but occupancy also increased for this site (~85% occupancy). In conclusion, the NxT modifications overall present a universal approach to increase PNGS occupancy on Env trimers but achieving full PNGS occupancy requires isolate-specific fine-tuning.

4.9 Discussion

This study sought to overcome the under-occupancy of PNGS on soluble HIV-1 Env SOSIP trimers from diverse virus isolates and establish guidelines for similar efforts on other vaccine antigens. The absence of a PNGS can easily be resolved by re-introducing the site, which has been successfully achieved for BG505 SOSIP trimers [282]. Under-occupancy of a PNGS is less easily addressed, but it is possible to increase occupancy on these trimers by manipulating the affinity of PNGS sequons for OST. In the final construct, 11 of the 12 NxS sequons were changed to NxT while, conversely, 1 of the 16 NxT sequons was switched to NxS. Nearly all of the PNGS were fully occupied on the resulting BG505 SOSIP NxT T135S T158S trimer, which now better resembled the viral BG505 trimer in this regard. *In vivo* studies should confirm whether elimination of artificial glycan holes alters immunogenicity. The N611 site appeared to be highly immunogenic for non-NAbs when under-occupied, judged by polyclonal-EM studies performed on sera from animals immunized with BG505 SOSIP.664 [299]. The modified SOSIP trimers should be a suitable design platform for further immuno-focusing efforts intended to facilitate the induction of bNAbs, although additional efforts should be made to close the large glycan hole present at the trimer base [300].

PNGS under-occupancy most likely occurs when the catalytic subunits of OST (STT3A and STT3B) skip sites, which occurs relatively frequently when the PNGS are close together (gap-0 or gap-1 sites) [276]. The WT protein was under-occupied at N133 (gap-1),

N190 (gap-0), N190c (gap-0), N197, and N611. In contrast, these same sites are fully occupied on viral Env [153]. Several factors associated with recombinant protein expression may be relevant to the differences. First, viral Env is tethered to the ER membrane via its transmembrane domain (in addition to the signal peptide) which might promote association with OST, particularly for PNGS near the C-terminus of the nascent polypeptide. Membrane tethering could also play a role in glycan processing, in particular those close to the C-terminus. However, other sites, such as N197, remain less processed on SOSIP trimers [301], and need other approaches to resolve these differences. The dwell time in the ER may also be affected resulting in more or less time for OST to act. Second, the tissue plasminogen activator (TPA) signal peptide used to express the WT protein, which is cleaved off co-translationally, might impact interactions with OST differently than for the WT HIV-1 signal peptide, which is cleaved off post-translationally [302,303]. Third, codon optimization of the recombinant SOSIP trimer may play a role via increased expression levels [304]. The much lower expression of viral Env compared to its recombinant counterpart could cause sequon skipping by OST because of low substrate abundance. Fourth, the use of different producer cell lines (HEK239F or CHO cells for SOSIP trimers, CD4+ T cells for BG505 virus may be relevant because of variation in their levels of OST and other components of the glycoprotein synthesis machinery.

NxT sequons are more efficiently glycosylated than NxS, probably because their affinity for OST is higher. The preference of OST for T over S at the +2 position of PNGS may arise from stronger van der Waals interactions [288,289]. Changing all 12 NxS sequons to NxT was compatible with efficient Env expression and trimer formation. When applied to the BG505 virus, infectivity was unaffected, a sensitive indicator that the 12 amino acid substitutions had no major adverse effects. PNGS occupancy on the NxT protein was increased at three sites surrounding the trimer apex, N190, N190c and N197, to the extent that they were now fully occupied (**Figure 4-4**).

In contrast to the increases seen for N190, N190c, N197 and N611, occupancy of N160 on the NxT protein was unexpectedly decreased, even though the N160 sequon itself was unchanged. The reduced occupancy of N160 diminished the binding of several V1/V2 bNAbs to the NxT protein because their epitopes involve this glycan. Neutralization of the NxT virus by the same bNAbs was also substantially reduced compared to the WT virus. There appear to be subtle and unpredictable influences on occupancy when PNGS are in proximity, as can be in the case in V1, for example at N133, and, presumably, other variable loops. There are reports that skipping by OST is likely to occur when one or two of the sequons for adjacent PNGS (gap-0 and gap-1) is NxS [276,286]. In contrast, attachment of glycans by OST is usually efficient when both sequons are NxT. As a result, double NxT sites are highly enriched in libraries of dually glycosylated gap-0 and gap-1 sites derived

from mammalian proteins. That finding is consistent with the observation that glycan attachment at the single gap-0 site in BG505 SOSIP trimers (N190 and N190c) was more efficient when these sequons were changed from NxS-NxS in the WT protein to NxT-NxT in its NxT counterpart.

The gap-1 sites N156-N160 and N133-N137 on the BG505 SOSIP trimer did not conform to the above pattern, which was unexpected. In both cases, occupancy was suboptimal when the two sequons were in NxT form, but there was an increase in occupancy when the first site was presented as NxS. One possible explanation is that the x-position of the N156 site is a cysteine that is involved in a disulphide bond, which could interfere with glycan attachment when OST affinity is increased by the NxT change. For both gap-1 sites a lower OST affinity of one of the sequons, whether driven by a lower on-rate and/or a higher off-rate, might allow a different OST complex to associate with the neighbouring sequon more efficiently. This kind of competition for OST between sequons might be much more pronounced in glycoproteins with an unusually high glycan density such as HIV-1 Env, compared to more typical human glycoproteins that carry only a few glycans. Hence, increasing glycan occupancy cannot be achieved by simply changing every sequon to NxT (although doing so often helps); the affinities of sequons for OST and how closely they are spaced are additional factors that need to be considered.

The Env trimer is the basis for HIV-1 vaccine research aimed at inducing broad and protective neutralizing antibodies. A now widely used immunogen design platform that is being explored in preclinical and clinical phase tests is the SOSIP trimer, a concept that has evolved from the initial prototype, BG505 SOSIP.664, to improved versions that better resemble viral Env [76]. SOSIP trimers can now be based on most viral sequences from various clades [219,281,291,292,305–309]. However, the SOSIP trimers evaluated to date do not fully mimic viral Env in respect of glycan occupancy; the resulting artificial holes in the glycan shield are immunogenic, but they induce antibodies of generally narrow specificity that are not on the pathway towards neutralization breadth. The NxT T135S T158S protein closely resembles viral Env in respect of both structure and PNGS occupancy. This next generation Env construct, or others based on the principles derived here, is a suitable starting point for additional immune-focussing efforts aimed at inducing glycan-dependent bNAbs. Finally, the PNGS sequon-engineering strategies described here could be applied to other highly glycosylated vaccine antigens and other biologics where PNGS occupancy could usefully be improved.

Chapter 5 Site-specific glycan analysis of the SARS-CoV-2 spike

5.1 Contributions

During the initial COVID lockdowns in April 2020, I worked on analysing the glycosylation of the SARS-CoV-2 spike with Dr Yasunori Watanabe, and we contributed equally across all areas of the study performed In addition, plasmid design and the production of one batch of SARS-CoV-2 spike protein was kindly provided by Dr Daniel Wrapp and Prof. Jason McLellan (University of Austin, Texas), who's stabilizing mutations allowed for the production of recombinant material for site-specific glycan analysis.

5.2 Introduction

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), the causative pathogen of COVID-19 [310,311], induces fever, severe respiratory illness and pneumonia. SARS-CoV-2 utilizes an extensively glycosylated spike (S) protein that protrudes from the viral surface to bind to angiotensin-converting enzyme 2 (ACE2) to mediate host-cell entry [43]. The S protein is trimeric class I fusion protein, composed of two functional subunits, responsible for receptor binding (S1 subunit) and membrane fusion (S2 subunit) [69,114]. Remarkably, the surface of the envelope spike is dominated by host-derived glycans with each trimer displaying 66 N-linked glycosylation sites. The S protein is a key target in vaccine design efforts [312], and understanding the glycosylation of recombinant viral spikes can reveal fundamental features of viral biology and guide vaccine design strategies [79,162].

Viral glycosylation has wide-ranging roles in viral pathobiology, including mediating protein folding and stability, and shaping viral tropism [61]. Glycosylation sites are under selective pressure as they facilitate immune evasion by shielding specific epitopes from antibody neutralization. However, as yet there have been no observed mutations to N-linked glycosylation sites [313]. Surfaces with an unusually high density of glycans can also enable immune recognition [61,314,315]. The role of glycosylation in camouflaging immunogenic protein epitopes has been studied for other coronaviruses [313,316,317]. Coronaviruses form virions by budding into the lumen of endoplasmic reticulum-Golgi intermediate compartments (ERGIC) [318,319]. However, observations of complex-type glycans on virally derived material suggests that the viral glycoproteins are subjected to Golgi-resident processing enzymes [316,320]. High viral glycan density and local protein architecture can sterically impair the glycan maturation pathway. Impaired glycan maturation resulting in the presence of oligomannose-type glycans can be a sensitive reporter of native-like protein

architecture [79], and site-specific glycan analysis can be used to compare different immunogens and monitor manufacturing processes [321]. Additionally, glycosylation can influence the trafficking of recombinant immunogen to germinal centres [322].

5.3 Site-specific glycan analysis of SARS-CoV-2 2P stabilized spike immunogen

To resolve the site-specific glycosylation of SARS-CoV-2 S protein and visualize the distribution of glycoforms across the protein surface, three biological replicates of recombinant soluble material were expressed and purified in an identical manner to that which was used to obtain the high-resolution cryo-electron microscopy (cryo-EM) structure, albeit without glycan processing blockade using kifunensine [69]. This variant of the S protein contains all 22 glycans on the SARS-CoV-2 S protein (**Figure 5-1A**). Stabilization of the trimeric prefusion structure was achieved using the "2P" stabilizing mutations [113] at residues 986 and 987, a "GSAS" substitution at the furin cleavage site (residues 682–685), and a C-terminal trimerization motif. This helps to maintain quaternary architecture during glycan processing. Prior to analysis, supernatant containing the recombinant SARS-CoV-2 S was purified by size-exclusion chromatography to ensure only native-like trimeric protein was analysed (**Figure 5-1B** and **Appendix F.1**). The trimeric conformation of the purified material was validated using negative-stain electron microscopy (**Figure 5-1C**).

To determine the site-specific glycosylation of SARS-CoV-2 S, trypsin, chymotrypsin, and alpha-lytic protease digests were performed to generate three glycopeptide samples. These proteases were selected to generate glycopeptides that contain a single N-linked glycan sequon. The glycopeptides were analysed by liquid-chromatography-mass spectrometry (LC-MS), and the glycan compositions were determined for all 22 N-linked glycan sites (**Figure 5-2**). To convey the main processing features at each site, the abundances of each glycan are summed into oligomannose-, hybrid- and categories of complex-type glycosylation based on branching and fucosylation. The detailed, expanded graphs showing the diverse range of glycan compositions is presented in **Appendix F.2** and **Appendix F.4**.

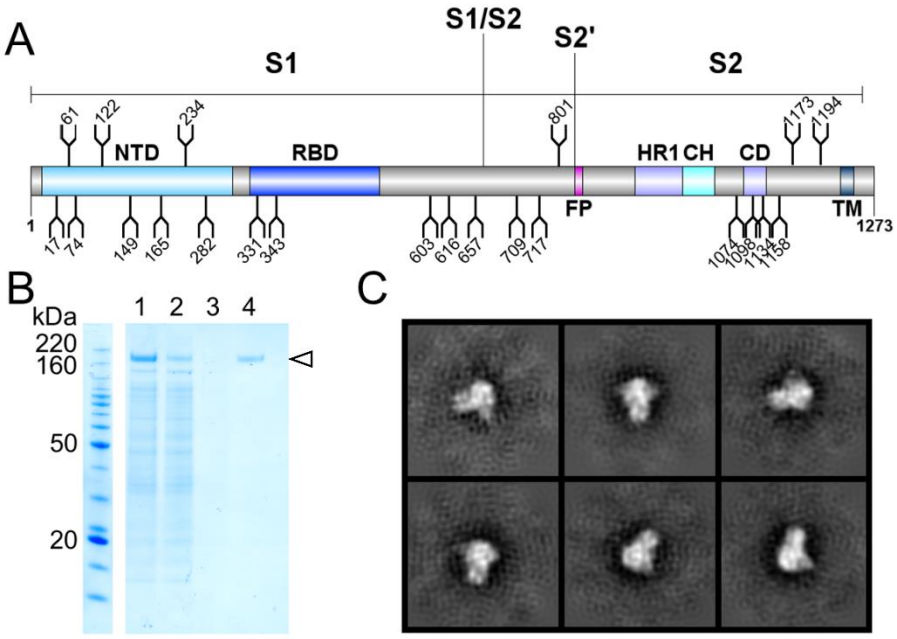


Figure 5-1: Expression and validation of SARS-CoV-2 S glycoprotein.
(A) Schematic representation of SARS-CoV-2 S glycoprotein. The positions of N-linked glycosylation sequons (N-X-S/T, where X≠P) are shown as branches. Protein domains are illustrated: N-terminal domain (NTD), receptor-binding domain (RBD), fusion peptide (FP), heptad repeat 1 (HR1), central helix (CH), connector domain (CD), and transmembrane domain (TM). (B) SDS-PAGE analysis of SARS-CoV-2 S protein expressed in human embryonic kidney 293F cells. Lane 1: filtered supernatant from transfected cells; lane 2: flow-through from StrepTactin resin; lane 3: wash from StrepTactin resin; lane 4: elution from StrepTactin resin. (C) Negative-stain EM 2D class averages of the SARS-CoV-2 S protein, performed by the laboratory of Prof. Jason McLellan, University of Austin, Texas. 2D class averages of the SARS-CoV-2 S protein are shown, confirming that the protein adopts the trimeric prefusion conformation matching the material used to determine the structure.

There are two sites on SARS-CoV-2 S that are principally oligomannose-type: N234 and N709. The predominant oligomannose-type glycan structure observed across the protein, with the exception of N234, is $Man_5GlcNAc_2$, which demonstrates that these sites are largely accessible to α1,2-mannosidases but are poor substrates for GlcNAcT-I, which is the gateway enzyme in the formation of hybrid- and complex-type glycans in the Golgi apparatus. The stage at which processing is impeded is a signature related to the density and presentation of glycans on the viral spike. For example, the more densely glycosylated spikes of HIV-1 Env and Lassa virus GPC exhibit numerous sites dominated by $Man_9GlcNAc_2$ [100,153,163,323].

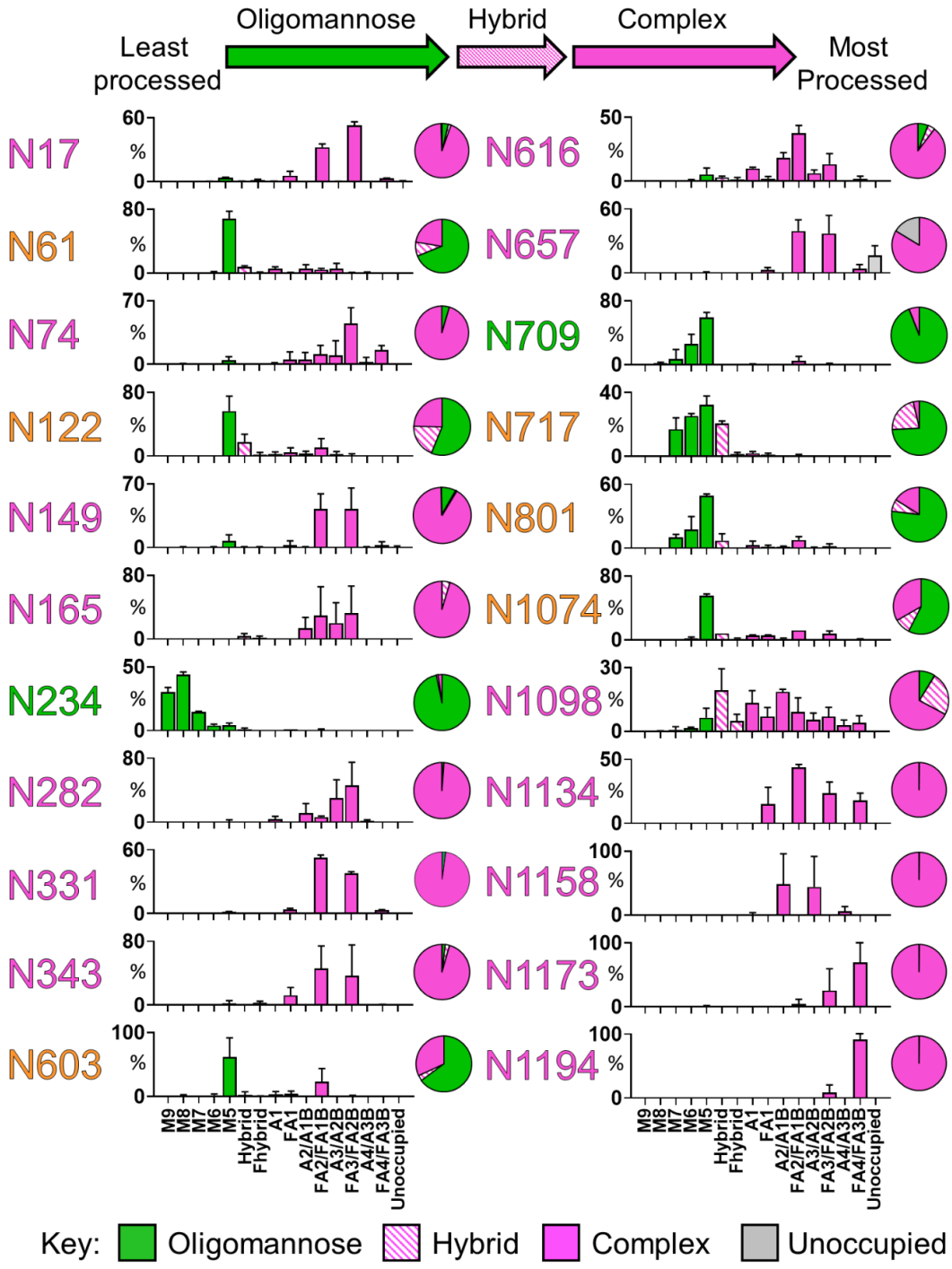


Figure 5-2: Site-specific N-linked glycosylation of SARS-CoV-2 S glycoprotein. The schematic illustrates the colour code for the principal glycan types that can arise along the maturation pathway from oligomannose-, hybrid- to complex-type glycans. The graphs summarize quantitative mass spectrometric analysis of the glycan population present at individual N-linked glycosylation sites simplified into categories of glycans. The oligomannose-type glycan series (M9 to M5; Man9GlcNAc2 to Man5GlcNAc2) is coloured green, afucosylated and fucosylated hybrid-type glycans (Hybrid & F Hybrid) dashed pink, and complex glycans grouped according to the number of antennae and presence of core fucosylation (A1 to FA4) and are coloured pink. Unoccupancy of an N-linked glycan site is represented in grey. The pie charts summarize the quantification of these glycans. Glycan sites are coloured according to oligomannose-type glycan content with the glycan sites labelled in green (80−100%), orange (30−79%) and pink (0−29%). An extended version of the site-specific analysis showing the heterogeneity within each category can be found in **Appendix F.4**. The bar graphs represent the mean quantities of three biological replicates with error bars representing the standard error of the mean.

A mixture of oligomannose- and complex-type glycans can be found at sites N61, N122, N603, N717, N801 and N1074 (**Figure 5-2**). Of the 22 sites on the S protein, 8 contain significant populations of oligomannose-type glycans, highlighting how the processing of the SARS-CoV-2 S glycans is divergent from host glycoproteins [324]. The remaining 14 sites are dominated by processed, complex-type glycans. Although unoccupied glycosylation sites were detected on SARS-CoV-2 S, when quantified they were revealed to form a very minor component of the total peptide pool (**Figure 5-2**). In HIV-1 immunogen research, the holes generated by unoccupied glycan sites have been shown to be immunogenic and potentially give rise to distracting epitopes [325]. The high occupancy of N-linked glycan sequons of SARS-CoV-2 S indicates that recombinant immunogens will not require further optimization to enhance site occupancy.

5.4 Contextualizing the site-specific glycosylation of spike protein using models

Using the cryo-EM structure of the trimeric SARS-CoV-2 S protein (PDB ID 6VSB) [69], the glycosylation status of the coronavirus spike mimetic was mapped onto the experimentally determined 3D structure (**Figure 5-3**). This combined mass spectrometric and cryo-EM analysis reveals how the N-linked glycans occlude distinct regions across the surface of the SARS-CoV-2 spike. Shielding of the receptor binding sites on the SARS-CoV-2 spike by proximal glycosylation sites (N165, N234, N343) can be observed, especially when the receptor binding domain is in the “down” conformation. The shielding of receptor binding sites by glycans is a common feature of viral glycoproteins, as observed on SARS-CoV-1 S [313,316], HIV-1 Env [326], influenza HA [327,328], and LASV GPC [323]. Given the functional constraints of receptor binding sites and the resulting low mutation rates of these residues, it is likely that there is selective pressure to utilize N-linked glycans to camouflage one of the most conserved and potentially vulnerable areas of their respective glycoproteins [252,329].

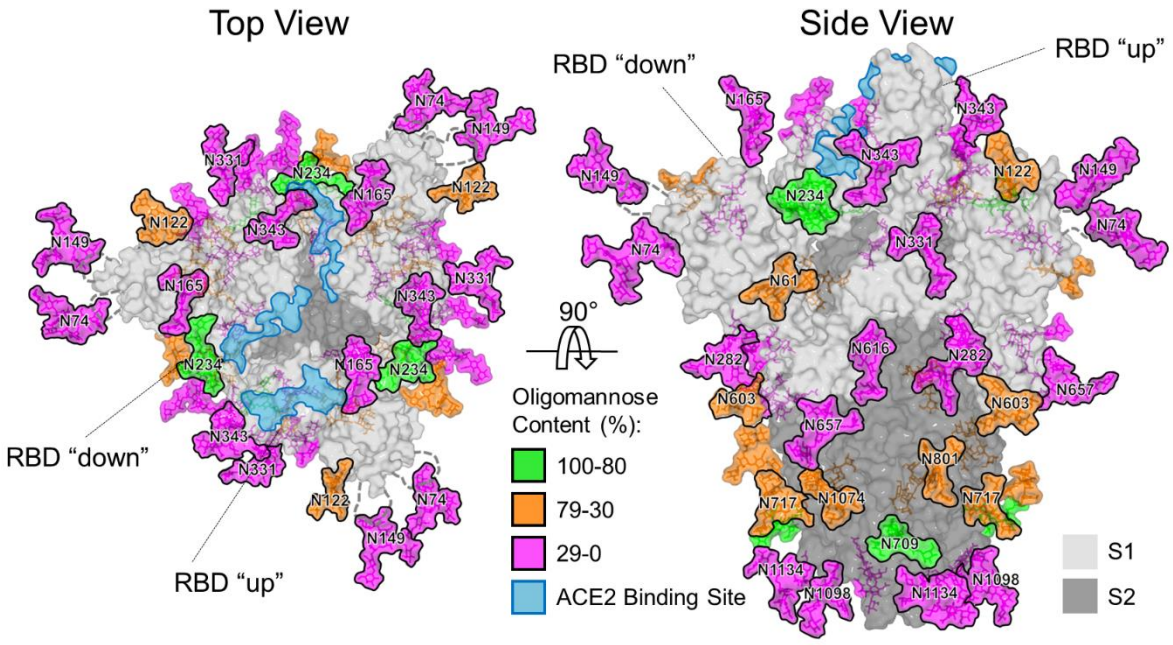


Figure 5-3: Structure-based mapping of SARS-CoV-2 S N-linked glycans.
Representative glycans are modelled onto the prefusion structure of trimeric SARS-CoV-2 S glycoprotein (PDB ID 6VSB) [69], with one RBD in the "up" conformation and the other two RBDs in the "down" conformation. The glycans are coloured according to oligomannose content as defined by the key. ACE2 receptor binding sites are highlighted in light blue. The S1 and S2 subunits are rendered with translucent surface representation, coloured light and dark grey, respectively. Note that the flexible loops on which N74 and N149 glycan sites reside are represented as dashed lines with glycan sites on the loops mapped at their approximate regions.

It is interesting to note the dispersion of oligomannose-type glycans across both the S1 and S2 subunits. This is in contrast to other viral glycoproteins, for example the dense glycan clusters in several strains of HIV-1 Env induce oligomannose-type glycans that are recognized by antibodies [80,330]. In SARS-CoV-2 S the oligomannose-type structures are likely protected by the protein component, as exemplified by the N234 glycan which is partially sandwiched between the N-terminal and receptor-binding (**Figure 5-3**).The N-linked glycans on extended flexible loop structures (N74 and N149) and at the membrane-proximal C-terminus (N1158, N1173, N1194) that were not resolved in the cryo-EM maps [69] were determined to be complex-type glycans, consistent with steric accessibility of these residues.

Whilst the oligomannose-type glycan content (28%) (**Appendix F.4**) is above that observed on typical host glycoproteins, it is lower than other viral glycoproteins. For example, one of the most densely glycosylated viral spike proteins is HIV-1 Env, which exhibits ~60% oligomannose-type glycans [153,264]. This suggests that SARS-CoV-2 S protein is less densely glycosylated and that the glycans form less of a shield compared with other viral glycoproteins including HIV-1 Env and LASV GPC, which may be beneficial for the elicitation of neutralizing antibodies.

Additionally, the processing of complex-type glycans is an important consideration in immunogen engineering, especially considering that epitopes of neutralizing antibodies against SARS-CoV-2 S can contain fucosylated glycans at N343 [331]. Across the 22 N-linked glycosylation sites, 52% are fucosylated and 15% of the glycans contain at least one sialic acid residue (**Appendix F.**4). This analysis reveals that N343 is highly fucosylated with 98% of detected glycans bearing fucose residues. Glycan modifications can be heavily influenced by the cellular expression system utilized. It has been previously demonstrated for HIV-1 Env glycosylation that the processing of complex-type glycans is driven by the producer cell but that the levels of oligomannose-type glycans were largely independent of the expression system and is much more closely related to the protein structure and glycan density [103].

Highly dense glycan shields, such as those observed on LASV GPC and HIV-1 Env, feature so-called mannose clusters [100,323] on the protein surface (**Figure 5-4**). Whilst small mannose-type clusters have been characterized on the S1 subunit of Middle East respiratory syndrome (MERS) CoV S [313], no such phenomenon has been observed for SARS-CoV-1 or SARS-CoV-2 S proteins. The site-specific glycosylation analysis reported here suggests that the glycan shield of SARS-CoV-2 S is consistent with other coronaviruses and similarly exhibits numerous vulnerabilities throughout the glycan shield [313]. Finally, trace levels of O-linked glycosylation were detected at T323/S325 with over 99% of these sites unmodified (**Appendix F.6**) suggesting that O-linked glycosylation of this region is minimal when the structure is native-like.

This glycosylation analysis of SARS-CoV-2 offers a detailed benchmark of site-specific glycan signatures characteristic of a natively folded trimeric spike. As an increasing number of glycoprotein-based vaccine candidates are being developed, their detailed glycan analysis offers a route for comparing immunogen integrity and will also be important to monitor as manufacturing processes are scaled for clinical use. Glycan profiling will therefore also be an important measure of antigen quality in the manufacture of serological testing kits. Finally, with the advent of nucleotide-based vaccines, it will be important to understand how those delivery mechanisms impact immunogen processing and presentation.

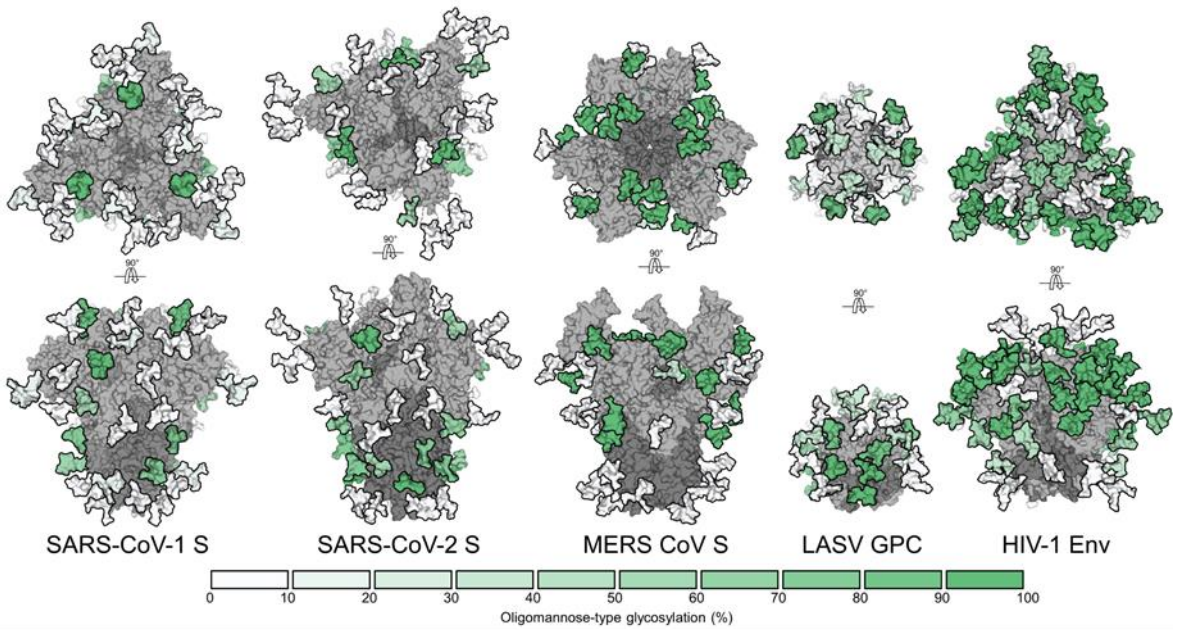


Figure 5-4: Under-processing of viral glycan shields.
Left to Right: MERS-CoV S [313], SARS-CoV-1 S [313], SARS-CoV-2 S, LASV GPC [323], HIV-1 Env [79,153]. Site-specific N-linked glycan oligomannose quantifications are coloured according to the key. All glycoproteins were expressed as soluble trimers in HEK 293F cells apart from LASV GPC, which was derived from virus-like particles from Madin-Darby canine kidney II cells.

Chapter 6 Site-specific steric control of SARS-CoV-2 spike glycosylation

6.1 Contributions

The goal of the work described in **Chapter 6** was to expand upon the work outlined in **Chapter 5** to include a comparison of SARS-CoV-2 S protein from multiple laboratories across the world. As such a range of collaborators contributed protein, the institutions which provided the proteins are listed in **Chapter 2.4**, with the CHO-derived spike protein provided by Excellgene and the monomeric RBD plasmids designed and expressed by the laboratory of Prof. Dennis Burton at the Scripps Research Institute. Additionally, LC-MS data for SARS-CoV-2 spike protein was prepared by Prof. Sai Li and Dr Yutong Song of Tsinghua University, which was reanalysed as part of this study to allow direct comparison. Molecular dynamics simulations to compare the glycan compositions to the solvent accessibility of each site were performed by the research group of Prof. Peter Bond A*STAR Bioinformatics Institute, Singapore. A full list of author contributions can be found in Allen *et al.* 2021 [332].

6.2 Introduction

The Coronavirus Disease 2019 (COVID-19) pandemic has prompted the development of an unprecedented array of vaccine candidates against the causative pathogen, severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2). All approaches aim to deliver molecular features of the virus in order to induce immunity. The viral spike glycoprotein, also referred to as S protein, has emerged as the principal focus of vaccine design efforts as antibodies against this target can offer robust immunity [194,201,214–216,333]. Encouragingly, neutralization can readily occur despite the extensive array of N-linked glycans distributed across the viral spike consistent with numerous vulnerabilities in this so-called glycan shield [334]. Despite these observations, glycosylation has emerged as an important parameter in vaccine development for SARS-CoV-1 and SARS-CoV-2 [335,336]. The glycosylation processing state can influence immunogen trafficking in the lymphatic system [322], influence the presentation of both native and unwanted cryptic epitopes [337], and reveal to what extent immunogens recapitulate native viral architecture. Evidence is also emerging that glycosylation can somewhat influence the interaction between SARS-CoV-2 and its target receptor, angiotensin converting enzyme 2 (ACE2) [120,338–340]. The SARS-CoV-2 S protein contains at least sixty-six N-linked glycosylation sequons that direct the attachment of host glycans to specific Asn residues. This extensive glycosylation is important in lectin-mediated protein folding and direct stabilization of the protein fold [341]. In addition, certain glycans are

incompletely matured during biogenesis and can lead to the presentation of immature glycans terminating with mannose residues that can act as ligands for innate immune recognition [103,118,323].

Despite the focus on the S protein in vaccine development efforts, there has been considerable divergence in the mechanisms of delivery. In one approach, nucleic acid encoding the spike is delivered through mRNA or with a viral-vector [210–217]. The resulting S protein is assembled and glycosylated by host tissue. In a contrasting approach, the S protein can be recombinantly manufactured either as recombinant protein using mammalian or insect cell lines or using inactivated virus-based approaches which allows detailed characterization of the immunogen prior to delivery [69,114,116,342,343]. Immunogen glycosylation can be influenced by factors specific to the manufacturing conditions such as cell type or cell culture conditions [344,345], however, construct design and protein architecture can also have substantial impact. For example, under-processed oligomannose-sites can occur at sites sterically hidden from the host mannosidase by the tertiary or quaternary architecture including obfuscation by neighbouring protein and glycan structure[118,346]. Immunogens displaying native-like architecture recapitulate these sites of oligomannose glycosylation. Conversely, immunogen design can adversely impact the presentation of native-like glycosylation. Importantly, despite the differences in biosynthesis of S protein in virions and from mammalian expression systems they seem to generate broadly similar glycosylation[228]. However, the success of a broad range of different vaccine platforms exhibiting different S protein glycosylation, indicate that native-like glycosylation is not a prerequisite for a successful vaccine. Despite this observation, understanding S protein glycosylation will help benchmark material employed in different serological and vaccine studies and help define the impact of this extensive feature of the protein surface.

The flexible and heterogeneous nature of N-glycosylation necessitates auxiliary methodologies in addition to cryo-electron microscopy or X-ray crystallography to characterize this key part of the S protein structure. Site-specific glycan analysis employing liquid chromatography-mass spectrometry is a widely used approach to obtain this information [61,100,154,347,348]. As research into the structure and function of the SARS-CoV-2 S protein has progressed, more details of the glycan shield of S protein have become apparent. Analyses on recombinant trimeric S protein revealed divergent N-linked glycosylation from host glycoproteins with the presence of under-processed oligomannose-type glycans at several sites [334,340,349]. Comparative analyses with monomeric and trimeric S proteins have revealed site-specific differences in glycosylation with regards to both oligomannose-type glycans and the presentation of sialic acid [349]. Analysis of S protein from insect cells demonstrated that oligomannose-type glycans were conserved on trimeric S protein, notably at N234 [350]. In addition, molecular dynamics (MD) studies have proposed

that the N234 site plays a role in stabilizing the receptor binding domain (RBD) in an exposed "up" conformation [346]. The presence of these under-processed oligomannose-type glycans on both mammalian and insect-derived S protein provides an indication that the structure of the S protein is driving the presentation of these glycans. Subsequent studies have investigated the presentation of N-linked glycans on S protein produced for vaccination, notably the Novavax full length S protein and S protein isolated following administration of the ChAdOx-nCoV-19 vector [350,351]. The observed glycan signatures were broadly in agreement with previous analyses. However, these studies involved the truncation of glycan structures using glycosidase treatment – which is useful for categorizing glycans into high mannose or complex-type glycans and determining potential N-glycosylation sites (PNGS) occupancy on low amounts of material – but does not allow for the identification of changes in terminal glycan processing such as sialylation. The glycan processing of the two N-linked glycan sites located on the RBD has also been investigated for monomeric RBD. These sites present high levels of complex-type glycans[352,353] and as the majority of antibodies raised against SARS-CoV-2 S protein target the RBD it is important to fully characterize the structure of the RBD, including the presentation of glycans.

N-linked glycans are highly dynamic and can substantially vary in chemical composition within a single batch of protein. Whilst glycosylation is heterogeneous, it is important for therapeutic and vaccine design to understand whether similar glycoforms arise across different protein expression platforms from different sources to ensure the antigenic surface of the S protein remains consistent when used as an immunogen or in serological assays [354–357].This is particularly important as glycan processing can be impacted by adverse protein conformations [79]. Here, an integrated approach including liquid chromatography- mass spectrometry (LC-MS) and MD simulations is applied to understand the glycosylation features across recombinant S proteins from different sources. This data was then compared to determine the extent to which these recombinant proteins reproduce the glycosylation of viral derived SARS-CoV-2 S protein produced from cultured Vero cells , obtained from a previous study [228]. An identical analytical approach was used to determine the glycan composition at each site, illustrating conservation across recombinant protein and virion derived material. Furthermore, the glycosylation of monomeric RBD recombinant protein was investigated, as it has previously been explored as a subunit vaccine and a candidate for serological testing [358,359]. The site-specific glycan analysis of the two sites located in the RBD of SARS-CoV-2, N331 and N343 when comparing monomer and trimer reveals a broad consensus in glycan processing, with some modest change in the processing of complex-type glycans. A similar comparative analysis of MERS-CoV RBD was performed to investigate whether the RBD glycans of MERS-CoV were under the same constraints as for SARS-CoV-2. The observed differences contrasted the glycan processing

of SARS-CoV-2 RBD monomer when compared with S protein, with trimeric MERS-CoV RBD glycan sites presenting restricted glycan processing, likely due to conformational masking of these sites, either by proximal glycans or nearby protein clashes, on trimeric MERS-CoV S protein [113,313]. The site-specific glycan data of SARS-CoV-2 S protein was then combined with MD simulations. These simulations reveal distinct degrees of accessibility between different glycan sites across the protein that broadly correlate to their processing states. Taken together, the results reveal the conserved structural N-glycan sites in S protein when compared with native viral spike which drives similarity in glycosylation amongst S protein from disparate sources and manufacture methods. Understanding S protein glycosylation will aid in the analysis of the vaccine and serological work of the global COVID-19 response.

6.3 Expression and purification of recombinant S protein from multiple sources

In order to define the variability in trimeric recombinant S protein glycosylation and compare recombinant and viral derived S protein we obtained preparations of recombinant S protein from a range of laboratories. These include the Amsterdam Medical Centre (Amsterdam), Harvard Medical School (Harvard), Excellgene (Switzerland), The Wellcome centre for Human genetics (Oxford) and Biological Sciences/University of Texas at Austin (Southampton/Texas) [69,194,226,227,232,360]. Whilst there are minor differences between the constructs used to produce the S protein the overall design is similar. This involves a truncation of the S protein prior to the C terminus, replacement of the furin cleavage site between S1 and S2 with a “GSAS” or otherwise mutated linker, a C-terminal T4 fibritin trimerization motif and stabilization of the prefusion S protein conformations using proline substitutions [69,113,361]. The recombinant proteins analysed contain 22 N-linked glycan sites with the exception of the “Amsterdam” preparation which was truncated prior to the last 3 N-glycans. An alignment of the recombinant proteins was performed and glycan sites are numbered according to the numbering used in previous reports of the site-specific glycosylation of the 2P stabilized S protein (**Appendix G.4**) [334]. All recombinant proteins were expressed in human embryonic kidney 293 (HEK293) cells with the Amsterdam, Harvard, and Southampton/Texas S protein preparations produced in HEK293F and the Oxford S protein prepared in HEK293T. The “Swiss” S protein was expressed in Chinese hamster ovary (CHO) cells. Following expression and purification, an identical approach was utilized to prepare the recombinant samples for analysis by LC-MS, involving three separate protease digests: trypsin, chymotrypsin, and alpha lytic protease. After analysis by LC-MS the site-specific glycosylation of these recombinant samples was compared to

previously published viral derived S protein which was searched using the same analytical parameters as for the recombinant proteins [228].

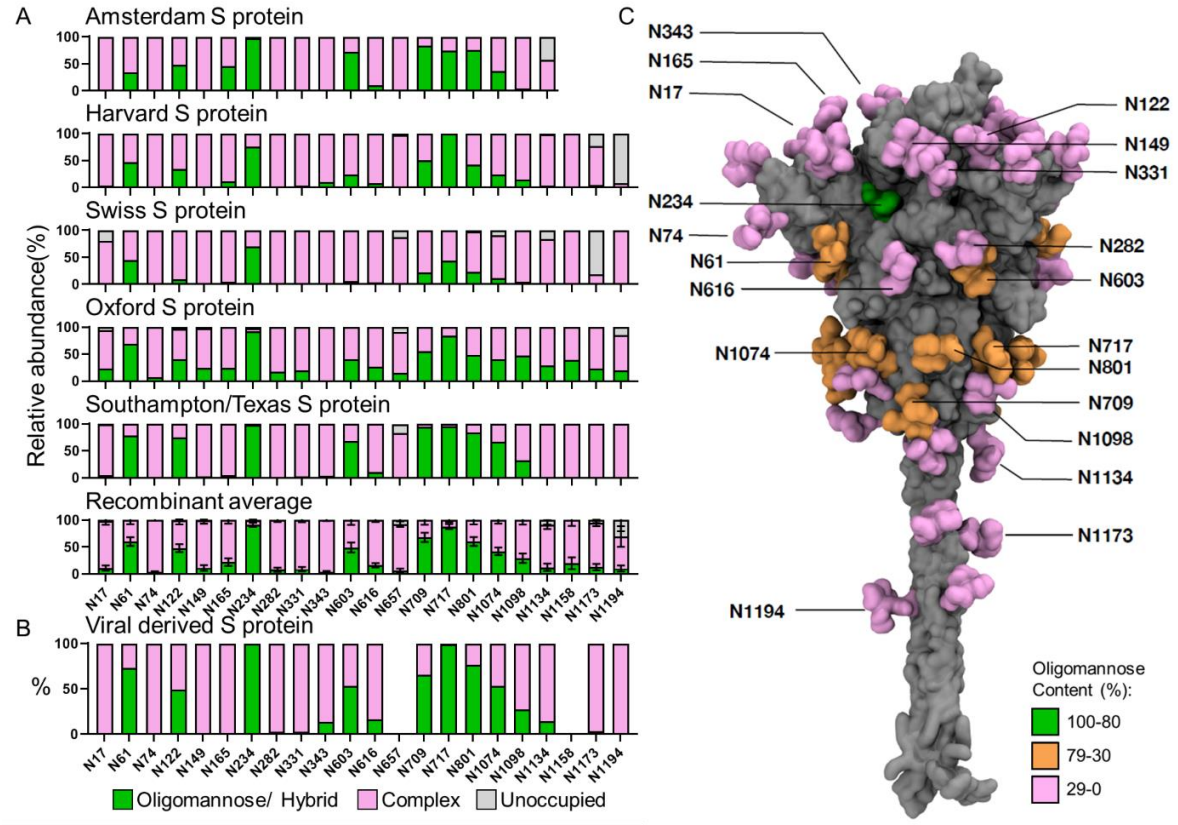


Figure 6-1: The site-specific glycosylation of recombinant and viral derived S protein from multiple laboratories.

(A) Site-specific glycan analysis of recombinant S protein expressed and purified at different locations. The bar charts represent the relative proportion of glycoforms present at each site, including the proportion of PNGS that were not modified by an N-linked glycan. The proportion of oligomannose- and hybrid-type glycans are coloured green. Processed complex-type glycans are coloured pink and the proportion of unoccupied sites are coloured grey. The institution which provided the S protein for analysis is listed above each chart. The Texas/Southampton data is reproduced from Chawla *et al.* (unpublished data). The average compositions of recombinant S protein were calculated using all samples. Bars represent the mean +/- standard error of the mean of all recombinant samples analysed. (B) The viral derived site-specific analysis was obtained from data acquired by Yao *et al.* and categorized in the same manner as above [228]. Data for sites N657 and N1158 could not be obtained and are not represented. (C) Full length model displaying the site-specific oligomannose glycosylation of viral derived S protein; description of how this model was generated can be found in Materials and Methods. Both protein and glycans are shown in surface representation; the former coloured in grey and the latter coloured based on oligomannose content as shown in Supplementary Table 1 (green for 80-100%, orange for 30-80% and pink for 0-30%).

6.4 Conservation of under-processed glycans on trimeric recombinant and viral-derived S protein.

Under-processed oligomannose-type and hybrid-type glycans are common on viral glycoproteins and commonly arise due to either glycan- or protein-mediated steric clashes with ER and Golgi resident mannosidase enzymes, terminating the glycan processing pathway. As the presence of these glycans is linked to the quaternary structure of the protein, changes in the abundance of these glycoforms can indicate changes in the fine structure of the glycoprotein. To investigate the abundance of these glycans, the heterogeneous glycan compositions detected by LC-MS were classified into three categories: (1) oligomannose-type and hybrid-type glycans, (2) complex-type glycans and (3) the proportion of PNGS lacking an N-linked glycan. Overall, the recombinant samples recapitulated the glycan processing observed on the viral derived S protein (**Figure 6-1A**). Sites on the viral-derived material with an abundance of under-processed glycans greater than 30% include N61, N122, N234, N603, N709, N717, N801 and N1074 (**Figure 6-1B**). With a few exceptions, all recombinant samples analysed also present at least 30% oligomannose-type glycans at these sites (**Appendix G.1**).

To compare the variability of oligomannose- and hybrid-type glycosylation across all recombinant proteins the compositions were averaged and displayed with the standard error of the mean (SEM) of S protein preparations from different laboratories (**Figure 1-9A** and **B**). This analysis revealed a broad consensus of glycan processing regarding high mannose glycans, with localized variations occurring. One example of remarkable homogeneity is the N234 site, which presents high levels of under-processed glycans in all samples analysed. This processing was conserved on the viral derived S protein. The recombinant material possessed low levels of oligomannose-type glycans at several sites across the protein which is not present on viral derived material. For example, material analysed from Oxford had at least 5% oligomannose-type glycans at every site except N343 (**Figure 6-1A**). This global moderate increase in oligomannose-type glycans could potentially arise from two sources, the first is that the recombinant preparations utilize stabilizing mutations that might generate a more compact structure which limits the ability of glycan processing enzymes to act, this is supported by the observation that viral-derived S protein exhibits metastable conformations in the NTD, likely increasing accessibility to particular glycan sites [228]. The second could be that the higher yield of protein obtained during recombinant protein expression places limitations on the capacity of the cell to process the high number of glycan sites present on the S protein.

Glycan occupancy is another important parameter to consider for immunogen design. Underoccupancy of recombinant glycoprotein PNGS compared to their viral counterparts

has been shown to result in the presentation of non-neutralizing distracting epitopes and is important to monitor [155,218]. As is the case with viral derived HIV-1 Env, viral derived SARS-CoV-2 S protein displayed high levels of glycan occupancy at all sites analysed (**Figure 6-1B**). When comparing this to the consensus site-specific glycosylation data, the majority of PNGS on recombinant S protein are highly occupied. The exceptions to this are at the C terminus, where several samples analysed display reduced glycan occupancy denoted by a larger population of unoccupied sites (**Figure 6-1A**). These sites displaying larger proportions of unoccupied glycans at the C terminus is likely due to the truncation of the recombinant proteins required for solubilization; it has previously been reported that the proximity of a glycan sequon to the C-terminus can influence its occupancy [362].

Modelling the site-specific glycosylation of viral derived material enables the glycosylation of S protein to be contextualized spatially in three dimensions. Using the cryo-EM structure of S protein ectodomain (ECD) in the open state (one RBD in the "up conformation" and two RBDs in the "down conformation") (PDB: 6VSB) [69], as well as the NMR structures of SARS-CoV HR2 domain (PDB: 2FXP) [230] and HIV-1 gp-41 TM domain (PDB: 5JYN) [231], a complete model of the S protein was generated (details in Methods section). This model includes glycosylation sites which are often not resolved as they are present on variable loops or along the flexible stem region of the S protein. This model demonstrated the spread of different glycan processing states across the S protein (**Figure 6-1C**). As with other published models of spike glycosylation, the N234 site is extensively buried within the protein surface proximal to the RBD and rationalizes the conserved presentation of oligomannose-type glycans on recombinant and viral-derived material as the glycan processing enzymes are unable to completely process these sites. Glycan sites located on variable loops and on the exposed stem of the S protein are much more processed as these sites are more exposed.

6.5 Divergent glycan processing at sites presenting complex-type glycans.

Whilst glycan under-processing can provide information pertaining to the quaternary structure of the glycoprotein, the majority of glycans across the S protein are able to be processed beyond these glycoforms. Glycan sites not under the same structural pressures as at the restricted sites such as N234 will be processed in an analogous manner as those presented on the majority of host glycoproteins. By comparing the site-specific glycan compositions at example sites, it is possible to investigate the variability in glycan elaboration across multiple preparations of S protein. To this end, the recombinant protein glycosylation data presented in this study (Amsterdam, Harvard, Swiss and Oxford) was compared with viral-derived S protein previously reported but analysed using the same

search parameters as for the recombinant proteins. To aid comparison, one site which presented high levels of under-processed oligomannose-type glycans (N234), one that presented a mixture of processing states (N1074), and a site that was populated by complex-type glycans (N282) was chosen for comparison (**Figure 6-2**). Comparing the site-specific glycosylation of N234 demonstrated the conservation of oligomannose-type glycan processing. On 4 of the 5 samples presented in **Figure 6-2** the most abundant glycan detected was $Man_8GlcNAc_2$, and on the Amsterdam S protein $Man_9GlcNAc_2$ predominated, which is marginally less processed. The Harvard and Swiss samples displayed elevated glycan processing at N234, but these glycans were only ~20% abundant and were still the least processed site on these samples (**Figure 6-2** and **Appendix G.1**).

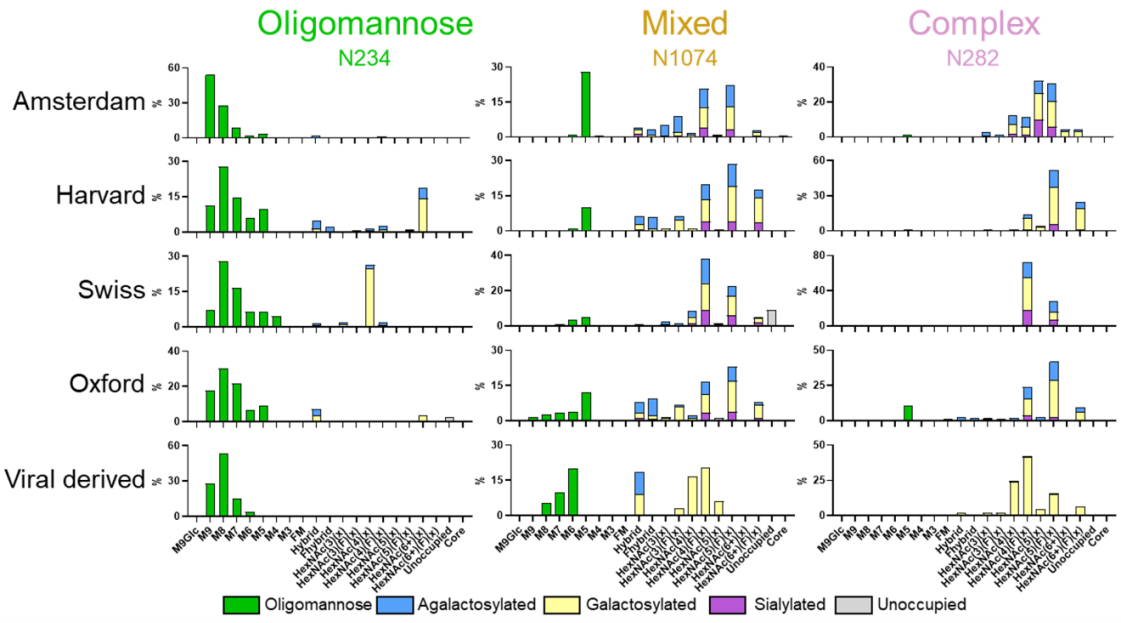


Figure 6-2 Detailed composition comparison between sites with differential processing states.
Glycan compositions at N234, N1074 and N282. For all samples analysed, glycans were categorized and coloured according to the detected compositions. Oligomannose-type glycans (M9-M4) were coloured green. Hybrid-type glycans, those containing 3 HexNAc and at least 5 hexoses, were coloured as for complex-type glycans since one arm can be processed in a similar manner. Complex-type glycans were categorized according to the number of HexNAc residues detected and the presence or absence of fucose. Core glycans represent any detected composition smaller than $HexNAc_2Hex_3$. For hybrid and complex-type glycans bars are coloured to represent the terminal processing present. Blue represents agalactosylated, yellow is galactosylated (containing at least one galactose) and purple is sialylated (containing at least one sialic acid). The proportion of unoccupied PNGS is coloured grey.

The N1074 site presented a diverse mixture of oligomannose-, hybrid-, and complex-type glycans across all samples analysed. The abundance of oligomannose-type glycans was more variable at this site with the total oligomannose-type glycan content varying from 11% to 35% (**Figure 6-2**). The predominant composition at N1074 varied from sample to

sample: for the Amsterdam recombinant preparation and the viral derived S protein, an oligomannose-type glycan was the most abundant glycan at N1074; however, for the other samples a fully processed glycan was more abundant. For all samples analysed, N1074 presented diverse glycan processing states. The elaboration of complex-type glycans with different monosaccharides can influence the function of the glycoprotein to which they are attached, for example, sialylation can extend the half-life of a glycoprotein in the body[121]. As they progress through the Golgi, glycans can be elaborated by the addition of fucose, galactose and sialic acid. The abundance of these monosaccharides is influenced by the cell from which they are produced and by the culture conditions or media compositions under which recombinant proteins are manufactured. Typically, glycoproteins produced from HEK293F and CHO cells present complex-type glycans with high levels of fucosylation and galactosylation but lower levels of sialylation [296,363]. The processing of N1074 and N282 demonstrates this with the majority of complex-type glycans bearing at least one fucose on all recombinant S preparations. Likewise, the majority of glycans are galactosylated, although populations of glycans are present which lack any elaboration beyond N-acetylglucosamine branching. Hybrid-type glycans, where one arm of the glycan is processed and elaborated as for complex-type glycans and one remains under-processed, are present in lower abundances across N1074 in recombinant preparations. These hybrid-type glycans are also generally of low abundance on mammalian glycoproteins [364].

Whilst the complex- and hybrid-type glycan compositions are variable, there are some visible trends when comparing glycan processing between recombinant and viral-derived S protein. The starkest difference is the lack of sialic acid residues across not only N1074 and N282 but across all PNGS of viral derived material (**Figure 6-2, Appendix G.1**). Upon averaging all the recombinantly produced S protein and comparing to viral derived S protein there is a 21 percentage-point decrease in sialylation (**Appendix G.2**). The fucosylation of viral derived material is also lower across both N1074 and N282 and this trend is again mirrored across all sites, with viral derived S protein possessing a 16 percentage-point decrease in fucosylation (**Appendix G.2**). The final difference is in glycan branching. The number of processed glycan antenna can be inferred from the detected number of N-Acetylhexosamine (HexNAc) residues determined by LC-MS. The more HexNAc present on a glycan composition the more branched that glycan tends to be, for example HexNAc(6) can correspond to a tetra-antennary glycan whereas HexNAc(4) is biantennary. As glycan processing is heterogeneous, this change is more subtle, however the complex-type glycans of N1074 and N282 are less branched on viral S protein compared to the recombinant S proteins, with a ~30 percentage point decrease in abundance of HexNAc(5) and HexNAc(6) corresponding to tri- and tetra-antennary glycans (**Figure 6-2**). These changes are also apparent on other processed sites such as N165 and N1158 (**Appendix G.2**).

These changes contrast expected observations based upon similar analyses comparing viral derived Env and soluble recombinant variants. For HIV-1 Env, an increase in glycan branching and sialylation is observed across sites presenting complex-type glycans [153]. These changes were also present on recombinant full length Env [301]. Several factors could be influencing these changes. The first is that the membrane tether afforded to the viral Env brings the glycans into close proximity to the glycan processing enzymes and the second is that the producer cell had greater expression levels of the glycosyltransferase enzymes involved in glycan processing. The SARS-CoV-2 S protein in a viral context is likely under similar constraints; however, the glycan processing is distinct. One factor which could be important is the early budding of the SARS-CoV-2 virion into the ER/Golgi intermediate complex (ERGIC) which may distance the spike protein from glycosyltransferases present in the trans-Golgi compared to Env, which remains attached to the membrane, proximal to glycosyltransferases [365]. The choice of producer cell and culture condition of the virus, in this case Vero cells, which are derived from the kidney cells of *Cercopithecus aethiops* (African green monkey), may also influence the glycan processing as the expression levels of glycosyltransferase enzymes may account for the diminished attachment of sialic acid and fucose observed on viral S protein compared to recombinant S protein. As these changes in glycosylation likely are not impacting the immunogenicity of the viral spike mimetics, as evidenced by the high efficacy of several vaccine candidates, these observations still remain important when considering how the virus may be interacting with the immune system via lectin interactions and may be informative when considering antigenic tests and purifications using glycan binding reagents.

6.6 The expression of monomeric RBD constructs impacts glycan processing

S protein glycosylation is influenced by quaternary protein architecture, and other factors. Next, the glycosylation of soluble recombinant monomeric RBD was compared with that of recombinant and viral derived trimeric S protein. Recombinant monomeric RBD was expressed and purified, to enable a comparison of the site-specific glycosylation of the two glycan sites located in the RBD, N331 and N343, with those observed on recombinant S protein and viral derived S protein reported previously[228,334]. Overall, the N331 and N343 glycans across all expression formats were highly processed, with little to no oligomannose-type glycans detected (**Figure 6-3**; and **Appendix G.3**). For the RBD sites presented on viral derived S protein, the complex-type glycans observed were similar to those observed at N282 with most of the PNGS occupied by bi- and tri-antennary glycans (HexNAc(4) and HexNAc(5)) (**Figure 6-3**). The majority of these glycans were fucosylated, however, a large

proportion of the complex type-glycans on viral derived S lacked fucosylation (24% N331 and 20% N343). As with other sites on the viral derived S protein, minimal levels of sialylation were observed and the majority of glycans possessed at least 1 galactose residue.

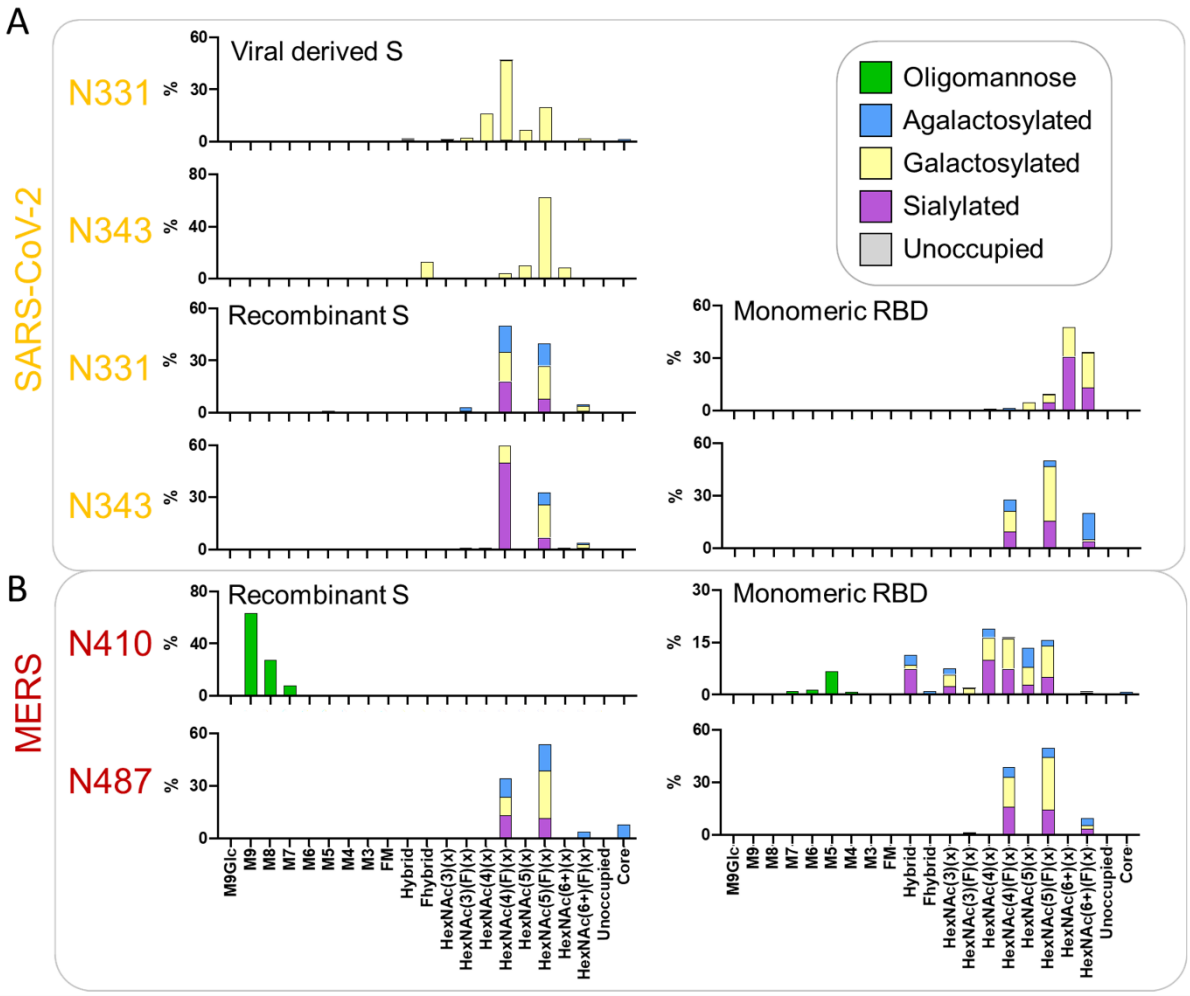


Figure 6-3 Comparative analysis of the glycosylation of the two PNGS on the RBD for viral derived S protein, recombinant S protein and monomeric RBD.
(A) Detailed site-specific glycan compositions of the two sites located in the RBD of SARS-CoV-2. Recombinant S protein data is reproduced from Chawla *et al.* (unpublished) and data for the viral S protein RBD sites was obtained from Yao *et al.* [228]. Site-specific glycan data is presented as outlined in **Figure 6-2**. (B) Site-specific compositions for N-glycan sites located in the RBD of MERS-CoV when expressed as part of a soluble recombinant S protein compared to RBD-only. Data for the MERS-CoV S protein were obtained from a previous study [313]. Site-specific glycan data is presented as outlined in **Figure 6-2**.

In comparison, the two RBD sites of recombinant soluble S protein are highly fucosylated, with close to 100% of the glycans at N331 and N343 containing at least one fucose (**Figure 6-3B**). As with the viral-derived S protein the majority of the glycans are bi- and tri-antennary complex-type glycans. The recombinant S protein RBD sites are also more sialylated than N331 and N343 on viral-derived S protein, with 28% of N331 glycans containing at least one sialic acid and 60% of those on N343 (**Appendix G.3**). Interestingly, when RBD is expressed as a monomer there are additional subtle changes compared to

viral derived and recombinant S protein. The most prominent change is in the glycan branching; whereas recombinant and viral derived RBD glycan sites possess low quantities of tetra-antennary glycans approximately, one third of the glycans at N331 and one fifth of the glycans at N343 consist of these larger branched structures (**Figure 6-3**). When compared to recombinant S protein the monomeric RBD sites also possess low levels of biantennary glycans. Despite these changes both recombinant trimeric S protein and monomeric RBD sites have high levels of fucosylation. These results suggest that complex-type glycans are under differential control hierarchies where certain forms of glycan processing could be influenced by the structural presentation of glycan sites. The attachment of fucose, galactose and sialic acid for these RBD sites appears to be controlled by more global phenomena, such as the producer cell, as the attachment of these monosaccharides is similar when comparing monomeric RBD and trimeric recombinant S protein which were produced in identical cell lines. The branching of the RBD sites on monomeric RBD is greater than that of trimeric S protein, both viral and recombinant, and suggests that the quaternary structure of the glycoprotein may have a small role on the elaboration of complex-type glycans of the RBD. These results are similar to previous analyses that have compared trimeric S protein with monomeric S1 [349].

To further explore differences in glycosylation of RBD PNGS, the glycosylation of MERS-CoV RBD was investigated. The site-specific glycan analysis of recombinant trimeric MERS-CoV S protein has been reported previously [313]. For the comparative analysis of trimeric recombinant protein and RBD, the MS files obtained in the previous study were searched using an identical version of the analysis software using the same glycan libraries as for the analysis displayed in **Figure 6-1** and **Figure 6-2**. This analysis revealed differences at the glycan sites present on MERS RBD when expressed monomerically compared to recombinant trimeric soluble S protein. One of the sites, N487, is similar between the two platforms, presenting glycoforms typical of sites populated by complex-type glycans (**Figure 6-3B**). In contrast N410 is occupied by exclusively oligomannose-type glycans when present on recombinant trimeric S protein. When monomeric MERS RBD is expressed the processing of the N410 site is markedly increased, with complex-type glycans predominating the glycan profile, although a subpopulation of oligomannose-type glycans remain (**Figure 6-3B**). Modelling the N410 glycan onto published structures of MERS S protein reveals that the N410 glycan is protected from processing by mannosidase enzymes such as ER α mannosidase I when buried within the trimer, but when presented on monomeric RBD it is readily accessible to glycan processing (**Appendix G.5**). These observations further highlight how the quaternary structure of a glycoprotein is a key determinant of the glycan processing state and demonstrates how glycan sites can provide information about protein folding and quaternary structure.

6.7 Molecular dynamics simulations reveal the relationship between accessibility and glycan processing.

Using site-specific glycan analysis, it is possible to infer the structural restraints placed on particular PNGS by the quaternary structure of the protein. MD simulations can help to understand how the protein flexibility can influence glycan processing. The recombinant proteins analysed in this study all use proline substitutions to stabilize the pre-fusion S protein conformation. Structural analyses of S protein containing these proline substitutions have shown that one of the three subunits frequently displays its RBD in an up conformation [69]. By performing simulations using models containing one RBD up, it is possible to investigate how differential RBD presentation can impact site-specific glycosylation.

To this end, 200 ns MD simulations of fully glycosylated trimeric S protein embedded within an ERGIC membrane model were performed (details in **Chapter 2.4.8**). The $Man_9GlcNAc_2$ glycan (Man-9), which represents the primary substrate for glycan processing enzymes, was added to each PNGS to understand the effect of protein and glycan dynamics on glycan processing. Glycan accessibility to enzymes was then elucidated by calculating the accessible surface area (ASA). To ensure the correct size of probe was used for ASA calculation, the structure of mannosidases and glycosyltransferases bound to its substrate or substrate analogue was used [164,366]. By measuring the distance between the substrate binding site and the outer surface of the enzymes, a probe with 1.25-1.5 nm radius would best approximate the size of the enzymes (**Appendix G.6**). ASA values for chain A of the S protein were measured using probes of 1.25 and 1.5 nm radii and found they both gave very similar results. Plotting accessible points of a 1.5 nm probe around individual glycans also indicated that a 1.5 nm radius would be required for the Man-9 glycan to be accessed by mannosidases and glycosyltransferases. Therefore, a 1.5 nm probe was used to measure ASA of each Man-9 glycan on the S protein model during the MD simulation.

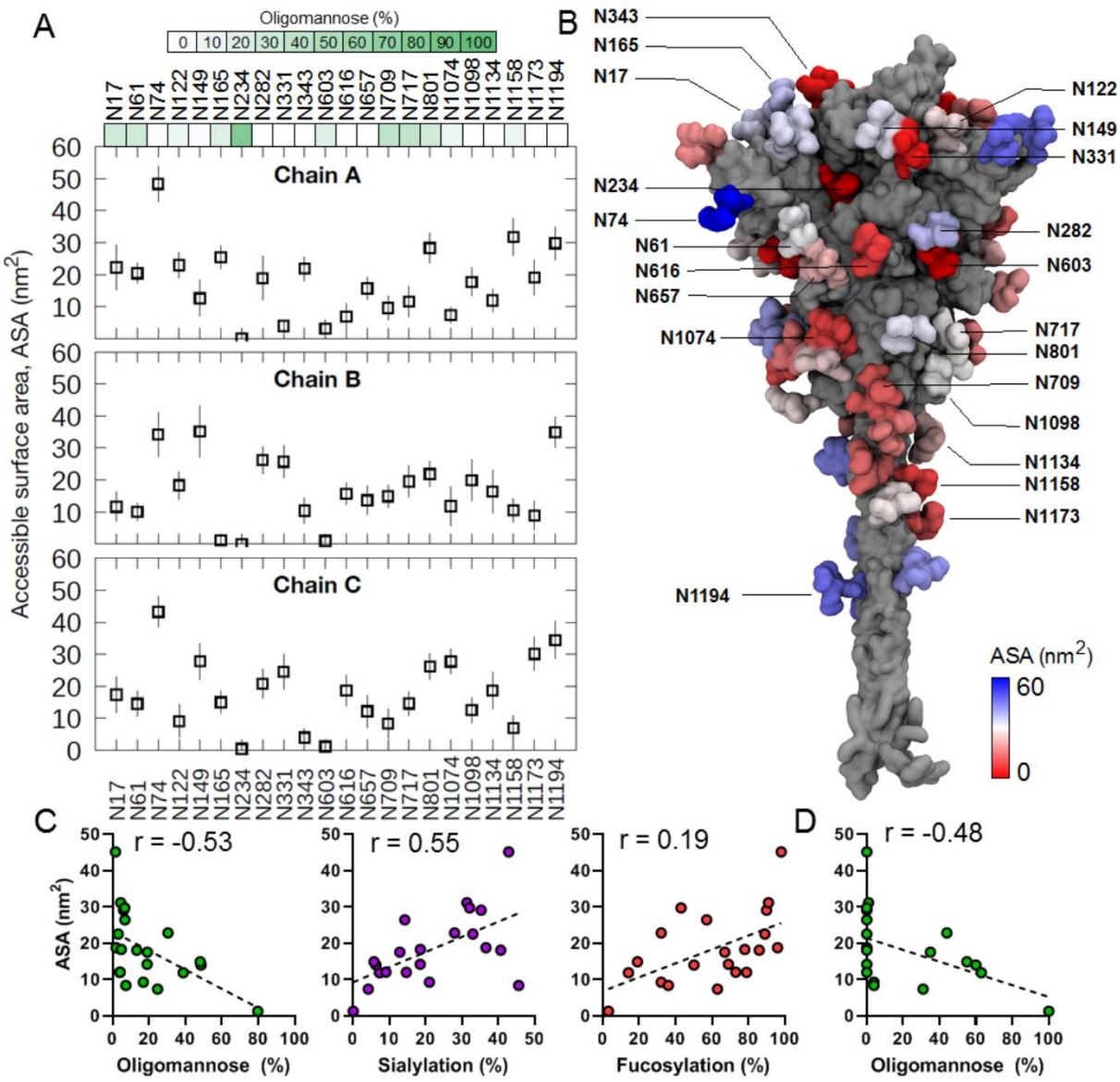


Figure 6-4 Accessible surface area (ASA) of oligomannose-type glycans from MD simulations.
(A) The ASA values were calculated for each oligomannose-type (M9) glycan for all three S protein chains. In this model, chain A is modelled in the RBD "up" conformation. The last 50 ns of the simulation was used for calculation and error bars indicate standard deviation along the trajectory. The probe size used was 1.5 nm. The green colour bars represent the oligomannose content of respective sites, calculated as the average oligomannose-type glycan content of the recombinant S proteins analysed in **Figure 6-1**(Amsterdam, Harvard, Swiss and Oxford S proteins). **(B)** The structure of the S protein (grey) with glycans coloured based on their ASA values shown in surface representation. **(C)** Graphs comparing glycan processing to calculated ASA values, averaged for chains A, B and C, for the recombinant protein average for the oligomannose-type glycan content, proportion of glycans containing at least one sialic acid and the proportion of glycans containing at least one fucose. The reported r values represent the Spearman's rank correlation coefficient. **(D)** The oligomannose-type glycan content versus ASA for viral derived SARS-CoV-2 S protein presented as in Panel C. The simulations detailed in panels A and B were performed by the laboratory of Prof. Bond, A*STAR Bioinformatics Institute, Singapore.

Comparing the ASA of each glycan across each protomer reveals a diverse range of accessibility (**Figure 6-4A** and **B**). Certain sites such as N234, were determined to possess low ASA values (**Appendix G.7A**), indicating that this site is extremely buried, correlating

with observations that N234 is the least processed on all recombinant and viral derived S protein. Conversely, the most exposed site across all three chains is N74 (**Appendix G.7B**). This glycan is highly processed when analysed by LC-MS. Interestingly, previous analyses have shown that N74 possesses sulphated glycans [340]. Sulphated compositions were included in the glycan library used to search the MS data and N74 contains multiple sulphated glycan compositions on all samples produced in HEK293 cells (**Appendix G.1**). The higher accessibility of this site may explain why sulphated glycans are observed at higher abundances at this site, but not others. Some glycans show a distinctly bimodal accessibility pattern, wherein high and low ASA values were observed along the trajectory (**Appendix G.8**). For example, two of the N122 glycans were buried within a crevice between the N-terminal domain of one chain and the RBD of an adjacent chain during a portion of the simulation. Similarly, one of the N603 glycans inserted into a large inter-protomeric groove near the S1/S2 cleavage site. This observation correlates with LC-MS data from both recombinant and viral derived S protein, showing that these two sites are populated by approximately half oligomannose/hybrid and half complex glycans, potentially due to the ability of these glycans to be either buried or exposed on the protein surface. The glycan on N165 also showed bimodal ASA values. However, this is due to the RBD up configuration, which allows N165 glycan to insert into the gap between the RBD and the N-terminal domain, which was not accessible in the RBD down configuration. N165, along with N234 glycans, have been shown to modulate the RBD's conformational dynamics by maintaining the up configuration necessary for ACE2 recognition [346].

To understand the relationship between oligomannose-type glycan presentation and the accessible surface area of each glycan site the average oligomannose-type glycan content of each site as determined in **Figure 6-1**, for both the recombinant and viral derived material, was compared to the average ASA values for chains A, B and C, as LC-MS cannot distinguish which glycan came from which chain (**Figure 6-4C** and **D**). For both recombinant and viral derived S protein, there was a negative correlation between ASA and oligomannose content (Spearman's correlation coefficient -0.53 for recombinant and -0.48 for viral derived). This demonstrates the link between glycan processing and quaternary architecture, as sites which are less accessible to the glycan processing enzymes present higher populations of immaturely processed oligomannose-type glycans. Conversely, the processing of complex-type glycans such as the addition of sialic acid correlated with the accessible surface area at a site-specific level ($r = 0.55$) and also fucosylation, albeit to a lesser extent ($r = 0.19$). Whilst there is an observed correlation between oligomannose-type glycans and ASA, the nanosecond time scale captured during MD simulations is shorter compared to the overall time taken for glycan processing in the ER and Golgi apparatus to occur, and therefore the S protein will be able to sample more conformations not resolved during simulation.

Finally, to determine the potential effect of glycan clustering on enzyme accessibility and subsequent glycan processing, the number of contacts made by each glycan at a given site with its surrounding glycans was determined (**Appendix G.9**). Interestingly, in all three chains, the sites on the S2 subunit show a higher number of glycan-glycan contacts, suggesting that they are most likely to form local clusters. An overlay of consecutive simulation snapshots of glycans also clearly shows that the S2 region is more densely packed with glycans compared to S1 and HR2 domains. This agrees with the LC-MS data showing that the sites on the S2 subunit, specifically N709, N717 and N801, have a high degree of unprocessed oligomannose-type glycans. The high density of glycan-glycan contacts within the mannose glycan patch has previously been implicated in reduced processivity by glycan processing enzymes in HIV-1 gp120 [146], and a similar mechanism might explain increased oligomannose content at S2 sites of SARS-CoV-2. A recent study showed that the glycan-binding anti-HIV-1 broadly neutralizing antibody 2G12 can bind to S protein via a common epitope around N709 on the S2 subunit [367], further suggesting the formation of a local glycan cluster. Collectively, MD simulation sampling shows that glycan accessibility to host enzymes, which is affected by protein quaternary structure and local glycan density, is an important determinant of glycan processing.

6.8 Perspectives

The global impact of COVID-19 has resulted in laboratories across the globe producing recombinant spike protein for vaccine design, antigenic testing, and structural characterization. Whilst the site of glycan attachment is encoded by the viral genome, the processing of the attached glycans can also be influenced by a wide range of exogenous phenomena, including recombinant host systems and processes for production in cell culture. Using immediately available materials, the glycosylation of Vero-produced virus preparations for Spike proteins was compared with those of HEK-293 and CHO-derived recombinant Spike proteins. These preparations are likely to be different from Spike proteins as produced in a diversity of cells in infected patients. However, when a glycan site is located in regions of the SARS-CoV-2 S protein which are not readily accessible the glycan site will possess high levels of under-processed oligomannose-type glycans, a phenomenon that will likely be of a general nature. This conclusion is derived from the observation that the under-processing of glycans is found in protein preparations from a range of institutions across the globe and are also present on Vero cell derived S protein derived from infectious SARS-CoV-2. Whilst the presence of oligomannose-type glycans are rare on the majority of host glycoproteins their abundance on SARS-CoV-2 is lower than that of other viral glycoproteins and means that the glycan shield density of this virus is low when compared to HIV-1 Env and Lassa GPC. This is likely to mean that the

immunodominant protein epitopes remain exposed. The processing equivalence between recombinant and viral derived S protein indicates that recombinant S protein glycosylation for vaccinations will likely mimic those occurring in human infections and will remain antigenically comparable. By analysing complex-type glycan processing across multiple samples, it has been shown that this glycosylation is driven more by other parameters including the producer cell and culturing conditions. One observed exception to this was that the expression of monomeric RBD as opposed to trimeric S protein does increase glycan branching on the two glycan sites located on SARS-CoV-2 RBD. This suggests that the glycan processing of complex-type glycans, in addition to oligomannose-type glycans, may also be under structural constraints, albeit to a much lesser extent. Overall, these results demonstrate that whilst N-linked glycosylation is highly diverse at certain regions of the S protein, there is a broad consensus of glycan processing with regards to oligomannose-type glycans between virus and immunogen S protein. This is something which cannot be taken for granted, as when comparing recombinant and viral derived HIV-1 Env, the reduced glycan occupancy of immunogens can induce an immune response that is incapable of neutralizing the virus. The reproducibility of S protein glycosylation from many different sources is of significant benefit for immunogen design, serology testing and drug discovery and will mean that the glycans of SARS-CoV-2 are unlikely to provide a barrier to combatting the COVID-19 pandemic.

Chapter 7 Subtle Influence of ACE2 Glycan Processing on SARS-CoV-2 Recognition

7.1 Introduction

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), the causative agent of COVID-19, encodes an extensively glycosylated spike (S) protein that protrudes from the viral surface [69,334]. The S protein is a trimeric class I fusion protein, composed of two functional subunits, S1 and S2, which are responsible for receptor binding and membrane fusion, respectively [69,114]. The S protein binds to cell surface angiotensin-converting enzyme 2 (ACE2) which initiates host-cell entry and results in membrane fusion [43]. Human ACE2 is a dimeric membrane-bound glycoprotein that encodes seven N-linked glycosylation sequons per protomer (UniProt KB: Q9BYF1). ACE2 glycosylation is likely to contribute to protein folding and stability [368,369], and evidence is emerging that it can influence SARS-CoV-2 recognition [370,371].

High resolution structures have been published of the SARS-CoV-2-ACE2 complex, however, one limitation of these structures is the difficulty resolving the contribution of glycan-glycan and glycan-protein interactions [372,373]. It is also known that glycans can influence the binding between two macromolecules even when glycans are attached outside of the protein-protein interface [374–376], and there is precedent for glycosylation of the host-receptor influencing viral binding interactions [377]. In the case of the SARS-CoV-2 interaction, removal of the ACE2 glycan at N90 enhances binding to the receptor binding domain (RBD) of the S protein and it has been suggested that the processing state of this glycan may influence the interaction [378]. The potential influence of ACE2 glycosylation on this interaction with the S protein has been further supported by molecular dynamics of the fully glycosylated complex [340,379]. It has been suggested that glycans may play a role in receptor binding as contacts are predicted between the glycans of ACE2 and both the protein and glycan components of the S protein [340,346,379]. For example, the glycan at N546 of the ACE2 receptor is suggested to interact with the glycans at N74 and N165 on the S protein [340]. Furthermore, the glycans of ACE2 at N90, N322 and N546 are all reported to interact with the protein moiety of the S protein [340]. In contrast, O-glycosylation abundance is low, with the exception of T730. However, this site is distal to the binding interface and likely does not contribute to binding [380]. Understanding the extent to which the glycan compositions, and the presence of the glycan itself, influence the binding of SARS-CoV-2

is important for both potential ACE2 therapeutic design and understanding the influence of host glycosylation on viral cell entry.

Beyond the steric effects of glycosylation on protein interactions, the binding capacity of viral glycoproteins directly towards carbohydrates can also be an important part of viral pathobiology. Indeed, the phenomenon of glycans on host-glycoproteins influencing and directly facilitating viral infection has been characterized on numerous viruses [61,348], with perhaps the most well-known example being the sialic-acid binding capabilities of hemagglutinin of influenza A viruses [61]. Importantly, a number of human coronaviruses have been shown to be able to bind sialosides, including MERS-CoV [119], SARS-CoV-1 [381], HKU1 [382] and SARS-CoV-2 [122,383,384]. The ability of SARS-CoV-2 to bind charged carbohydrates is further supported by the observation that sulphated glycosaminoglycans bind to the viral spike glycoprotein and can inhibit infection [385–389]. With the capacity of SARS-CoV-2 to directly recognize sialylated glycans it provides further impetus to explore the impact of ACE2 glycosylation on the binding to the S protein.

Here, the glycan structures on ACE2 were modified using several glycan engineering techniques, which incorporated the use of glycosyltransferases, glycosidases, and enzymatic inhibitors. Using a mass spectrometric approach, the site-specific glycan structures present on ACE2 were determined both with and without glycan engineering. Binding studies of the glycan variants of ACE2 were performed using surface plasmon resonance (SPR) to determine binding affinities between SARS-CoV-2 S and ACE2 variants. These experiments demonstrated that when ACE2 glycans are hypersialylated, or when all glycans were converted to oligomannose-type, there was a modest decrease in affinity. When the sialic acid residues were removed a significant but modest increase in affinity was observed. However, overall, the effects were subtle and deglycosylation had minimal impact on S binding.

7.2 Expression and purification of ACE2

To enable the determination of the glycosylation and binding capabilities of ACE2, a soluble recombinant ACE2 ectodomain that possesses all, but the last N-linked glycan site was expressed in HEK 293F cells (**Figure 7-1A**). This monomeric material has previously been shown to bind S protein and has been structurally characterized by cryo-electron microscopy [69]. Expressed material was purified by nickel-affinity chromatography and size exclusion chromatography (SEC). The SEC chromatogram demonstrated a single peak (**Figure 7-1B**) which exhibited a mass of ~75 kDa when subjected to SDS-PAGE analysis (**Figure 7-1C**). The soluble ACE2 expressed at ~80 mg/mL and ensured that all experiments performed were undertaken on the same batch of ACE2. To assay ACE2 binding SARS-CoV-2 S protein was expressed and purified in an identical manner to ACE2, as reported previously [69,334].

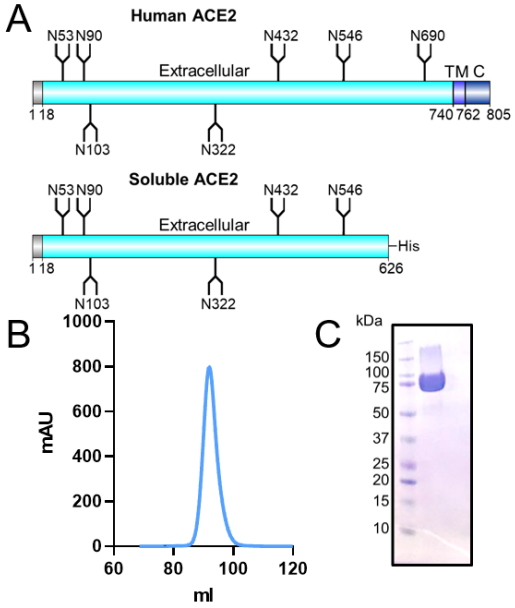


Figure 7-1: Production and purification of soluble ACE2.
(A) Schematic of ACE2 comparing full length human ACE2 with the soluble ACE2 construct used in this study shown underneath. The positions of N-linked glycan sequons (N-X-S/T, where X≠P) are shown as forked sticks. (B) Size exclusion chromatogram of soluble ACE2 following nickel affinity purification. (C) SDS-PAGE using Coomassie stain of the pooled SEC fractions for soluble ACE2.

7.3 ACE2 glycosylation analysis and SARS-CoV-2 S binding

To understand the impact of the ACE2 glycosylation upon SARS-CoV-2 binding the glycans present on the soluble unmodified variant used in this study were determined (**Figure 7-1A**). Human ACE2 encodes seven N-linked glycosylation sites at N53, N90,

N103, N322, N432, N546 and N690. The soluble construct used here lacks the N690 glycan, however, this glycosylation site lies distal to the protein interface with the S protein and is therefore not predicted to directly influence the interaction [340]. To resolve the site-specific glycosylation, trypsin, chymotrypsin, and alpha-lytic protease was used to generate glycopeptides, which were subsequently subjected to liquid-chromatography-mass spectrometry (LC-MS) analysis (**Figure 7-2A**). This analysis revealed a high abundance of complex-type glycans on ACE2, as would be expected for a mature, secreted mammalian glycoprotein. These observations are similar to those previously reported for various recombinant ACE2 glycoproteins [340,380]. Interestingly, whilst the glycan sites at N53, N90, N103, and N546 were occupied with complex-type glycans, N322 and N432 exhibited low levels of glycan site occupancy with the unmodified peptide accounting for 72% and 62% for each respective site. This underoccupancy has been observed in another analysis of ACE2 glycoproteins [340]. The extent of under occupancy may exceed that of native, full-length ACE2 and be driven, to some extent, by the high expression levels of ACE2 in the recombinant system. This phenomenon has been observed for HIV-1 where soluble immunogens exhibit lower glycan occupancy compared to full-length, membrane bound, material [153,155].

As an LC-MS methodology analysing intact glycopeptides was used it was therefore possible to categorize the heterogenous glycan populations according to their terminal monosaccharides (**Figure 7-2A**). Glycans were grouped according to the number of processed antennae, the presence or absence of fucose, the presence or absence of galactose and if the glycan contains sialic acid. The predominant type of glycan observed across all five sites were fucosylated bi- and tri-antennary complex-type glycans which lack galactose and sialic acid. The exception to this is at N103 where a high abundance of sialylated glycans was observed. Overall high levels of fucosylation were observed across all glycans (66%; **Appendix H.1**). The galactosylation was moderately abundant (37%) and sialylated N-glycans were comparably low (15%). This pattern of high fucosylation, moderate galactosylation and low sialylation is typical of protein expressed using recombinant cell lines such as HEK 293F [109,153,390].

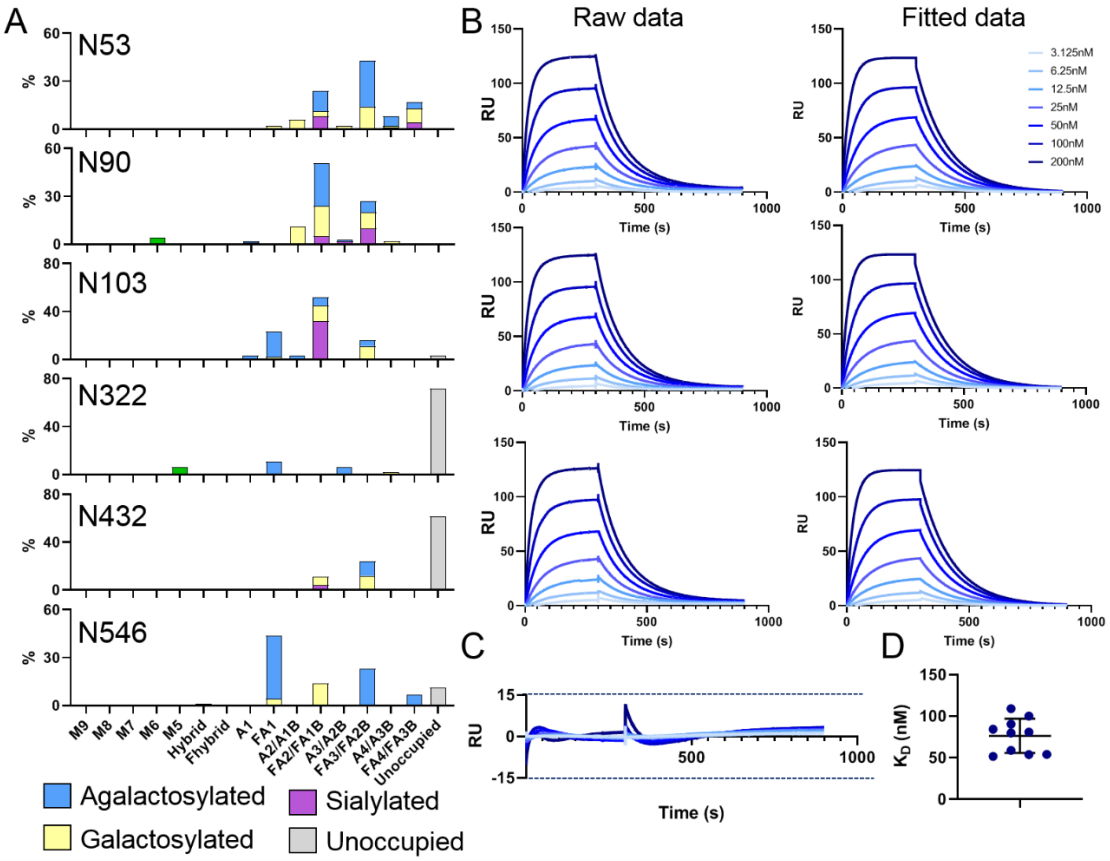


Figure 7-2 Site-specific glycan analysis and determination of SARS-CoV-2 binding of soluble ACE2.
(A) Site-specific glycan analysis of soluble ACE2 determined by LC-MS. Glycans are categorized into oligomannose-type glycans (green and Complex-type glycans which are grouped according to the presence/absence of fucosylation as well as the number of processed antennae. Within these categories glycans are coloured according to the presence of galactose (yellow), sialic acid (purple), and the absence of both (blue). The proportion of N-glycan sites which do not have a glycan attached are coloured grey. Nomenclature: M5-M9, $Man_{5-9}GlcNAc_2$; F, fucose; A, antennae; B, bisecting GlcNAc. (B) Surface plasmon resonance sensograms for soluble ACE2 binding to SARS-CoV-2 S protein. One of the ten soluble ACE2 SPR experiments is shown. Each sensogram represents one analytical repeat. The fitted were generated from the three repeats using Biacore Evaluation software with a 1:1 binding model. (C) Residuals comparing the deviation of the fitted data to the experimental data. (D) Plot of the K_D determined for each experimental repeat of soluble ACE2 binding to SARS-CoV-2 S protein. The mean is plotted as a black line and the error bars represent +/- standard deviation as calculated using GraphPad Prism.

To understand the impact of different glycans upon ACE2 binding, the interaction between the un-engineered soluble ACE2 (referred to here as the wild-type form, (WT) and SARS-CoV-2 S glycoprotein was determined. SPR was used to define the dissociation constant (K_D). Prior to analysis, the His-Tag at the C-terminus of ACE2 was removed using HRV3C protease to minimize ACE2 binding to the sensor chip. For each analysis three repeats were performed with serial dilutions of ACE2 ranging from 200 nM to 3.125 nM (**Figure 7-2B**). To calculate the binding parameters a fitted curve was generated using a 1:1 binding model. The sensograms showed minimal deviation when compared to the fitted curves (**Figure 7-2C**). When calculating binding kinetics, the average of three analytical

repeats was used. To understand the batch-to-batch variation in SARS-CoV-2 affinity for ACE2, a series of experiments were performed with different batches of SARS-CoV-2 glycoprotein with a single ACE2 control in each in which no glycan engineering had been performed. In total, 10 ACE2 sensograms were obtained. The measured K_D for the interaction between SARS-CoV-2 and ACE2 varied between 50 and 100 nM with an average of 76 nM (**Figure 7-2D**).

7.4 ACE2 glycosylation modestly impacts binding to SARS-CoV-2

In order to probe the influence of different glycoforms of ACE2 on SARS-CoV-2 S binding, glycan modified versions of the receptor were generated. A range of different approaches were used including modifying the glycans during expression of ACE2 and using glycosidases post-expression to remove specific monosaccharides. To investigate the impact of sialylation upon ACE2 binding, ACE2 was co-expressed with a solubilized variant of β-galactoside α-2,6-sialyltransferase I (ST6) [391]. As the ST6 contains a His-Tag, the SEC contained two peaks which, when analysed by SDS-PAGE, confirmed two proteins, one at ~75 kDa and one at 35 kDa, corresponding to ACE2 and ST6 respectively (**Figure 7-3A**). To determine whether ST6 co-expression increased sialylation the same LC-MS approach was used, as described for WT ACE2. This demonstrated a notable increase in sialylation at every site, with the exception of N103 which exhibits a high degree of sialylation on the WT protein and was not enhanced by ST6 co-expression (**Figure 7-3B**). The total sialylation increased from 15% for WT ACE2 to 54% for ACE2 co-expressed with ST6 (**Appendix H.2**), noting that the overall levels are somewhat suppressed by the population of aglycosylated peptides. The levels of fucosylation and unoccupancy were comparable to the WT ACE2.

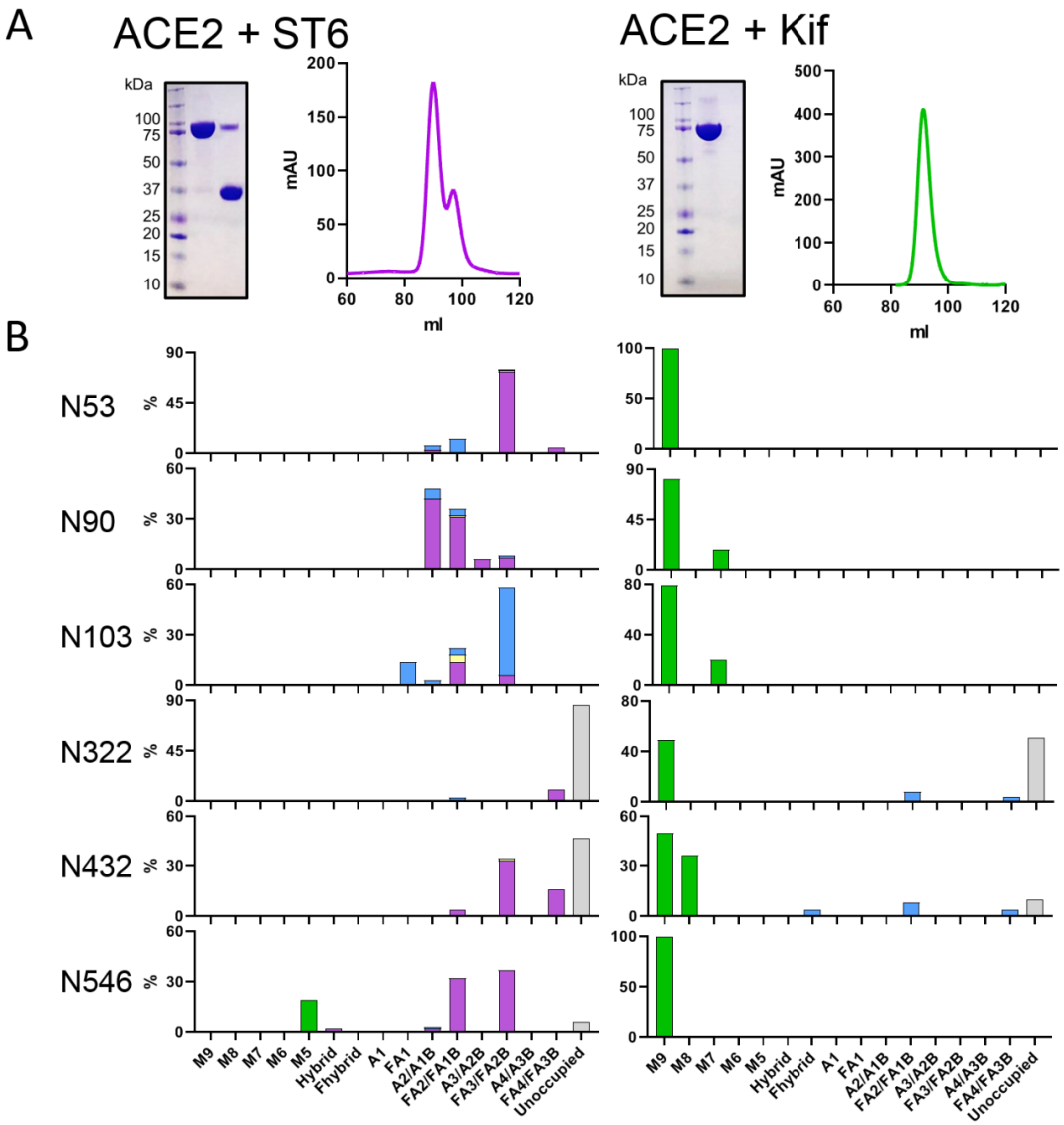


Figure 7-3 Glycan modification of soluble ACE2 using ST6 and kifunensine co-expression.
(A) SDS-PAGE and SEC trace for ACE2 co-expressed with ST6 and also ACE2 expressed with 20 µM kifunensine. The two peaks shown in SEC were pooled separately and show two distinct bands corresponding to soluble ACE2 (~75 kDa) and ST6 (~37 kDa). **(B)** Site-specific glycan analysis of ACE2 expressed with ST6 and ACE2 expressed with kifunensine. Bars are categorized and coloured according to **Figure 7-2**.

In addition to sialylation, the impact of artificially adding oligomannose-type glycans at each site across ACE2 was investigated. These glycans are rare on healthy host glycoproteins and it is unlikely these glycans are in high abundance on cellular ACE2, although oligomannose-type glycans can arise naturally on cellular receptors, depending on the producer cell [324,392]. To convert the glycans to predominantly $Man_9GlcNAc_2$, 20 µM of kifunensine (Kif) was added at the time of transfection. This enzyme inhibits the glycan-processing enzyme, ER α-mannosidase I, immediately following protein folding which results in the presentation of oligomannose glycans on the mature glycoprotein. The resultant ACE2, which is referred to as Kif ACE2, was purified and, as for WT ACE2, a single peak was observed by SEC (**Figure 7-3A**). Glycan analysis confirmed inhibition of

glycan processing, with large populations of oligomannose-type glycans at every site (**Figure 7-3B**).

To further modify the glycans of ACE2 the three glycan variants were used as a platform to generate several glycan modified variants of ACE2. For each approach LC-MS was performed to confirm successful enzymatic digestion of the glycans (**Appendix H.3**). To assay whether the presence of a glycan impacts ACE2 binding the oligomannose-type glycans were removed from Kif ACE2 using endoglycosidase H (endoH) which leaves a single *N*-acetylglucosamine residue [393]. This approach was taken as opposed to removing all the glycans from WT ACE2 as the total removal of N-glycans from proteins can cause misfolding, whereas the contribution of a single *N*-acetylglucosamine to ACE2 binding will likely be minimal. This approach is commonly utilized to obtain crystal structures of deglycosylated proteins while avoiding the complications of misfolding and incomplete digestion that can be associated with the use of peptide N-glycosidase [110,393]. To further understand the impact of sialylation, the sialic acids were removed from both WT and ST6 ACE2 using α2-3,6,8 sialidase. The role of fucose was investigated using WT ACE2 treated with α-fucosidase. Each of these proteins were characterized by LC-MS and used for subsequent analysis by SPR. Overall, it is possible to modulate the glycosylation of ACE2 in a specific manner for downstream analyses.

Using an identical experimental set up as for WT ACE2, a series of experiments were performed using both WT ACE2 and glycan modified ACE2. When ACE2 was co-expressed with ST6 a modest increase in K_D was observed, which corresponds to a decrease in binding (**Figure 7-4A**). When terminal sialic acid residues are removed using sialidase from both WT and hypersialylated ACE2 the affinity does not revert to WT but instead converge to a K_D ~30% higher, indicating that ACE2 treated with sialidase has a higher affinity for SARS-CoV-2 compared to WT ACE2 (**Figure 7-4B** and **F**). The observation that the viral S protein interaction with ACE2 is only slightly influenced by ACE2 sialylation suggest that this interaction is robust to variations in sialylation state. Similarly, the presence or absence of fucose does not impact ACE2 binding to a significant extent (**Figure 7-4E**). When all N-linked glycans on ACE2 are converted to oligomannose-type glycans using kifunensine the K_D increases by ~50%, corresponding to a decrease in binding (**Figure 7-4C**). However, when these oligomannose-type glycans are removed by endoH, leaving only a single N-acetylglucosamine residue remains, the affinities are comparable with WT ACE2 (**Figure 7-4D**). However, the impact of mannosylation suggest that viral S protein more optimally binds ACE2 glycoprotein that has undergone glycan maturation during biogenesis.

These effects are subtle, and it is an important observation that deglycosylated ACE2 remains capable of binding SARS-CoV-2 S protein (**Figure 7-4D**). To confirm initial observations regarding the binding of glycan modified ACE2 to SARS-CoV-2 the results

were reproduced and confirmed the initial observations. These experiments confirmed the significant difference between untreated ACE2 and sialidase treated ACE2 (p = 0.04) and also ST6 ACE2 treated with sialidase (p = 0.03) (**Figure 7-4F** and **Appendix H.4**) but also highlighted the subtlety of the observed effects as no other observed differences were significant. There was extensive variability between ST6 ACE2 experiments and the observed differences in K_D were not statistically significant. This could be due to the impact of microheterogeneity amongst hypersialylated ACE2 glycoforms leading to elevation in variability when compared to other samples.

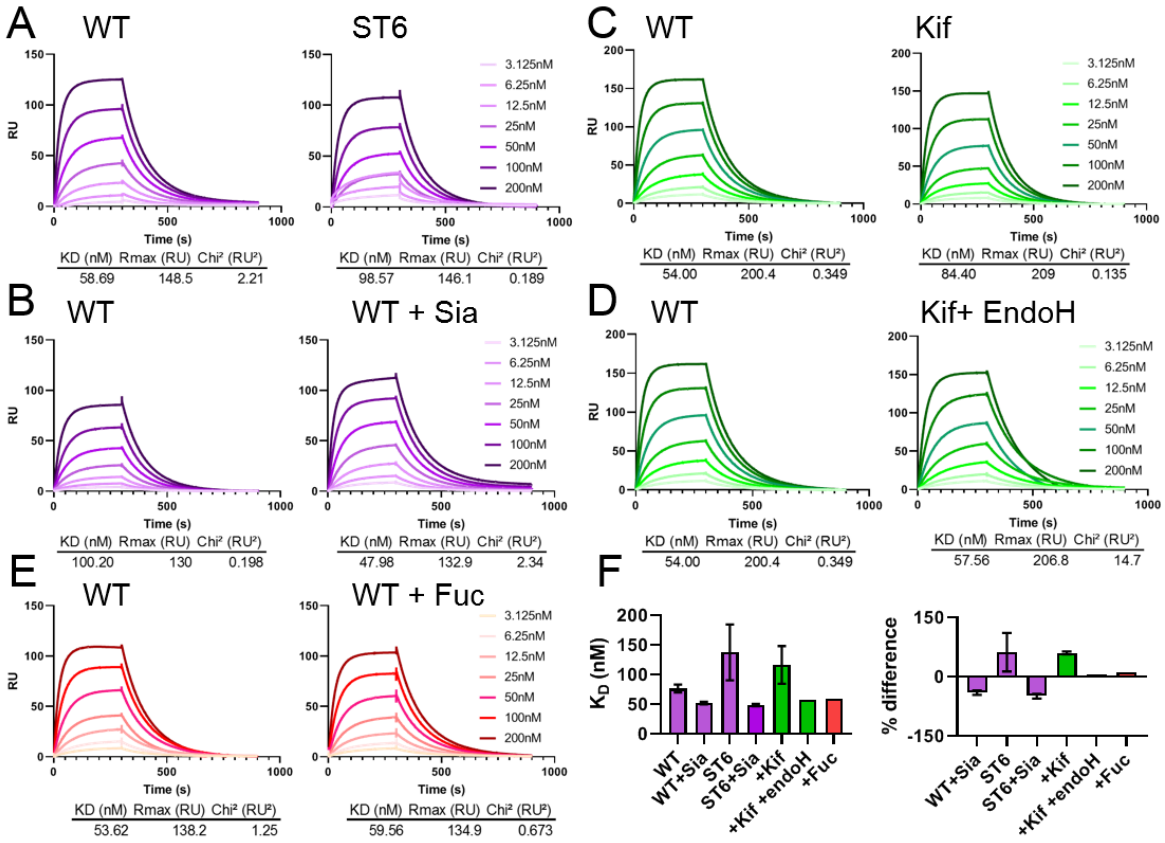


Figure 7-4: SPR analysis of glycan modified soluble ACE2 binding to SARS-CoV-2 S protein.

(A) Comparison of sensograms for WT ACE2 and ACE2 co-expressed with ST6. Both sensograms were obtained using the same SARS-CoV-2 S protein preparation, each WT ACE2 experiment detailed was performed using the same batch as the corresponding glycan modified ACE2. All displayed curves represent the average of three analytical repeats. All kinetic parameters were calculated using a 1:1 binding model using the Biacore evaluation software. (B) Sensograms comparing WT soluble ACE2 and WT ACE2 treated with sialidase. (C) Sensograms comparing WT ACE2 and ACE2 expressed in the presence of 20 µM kifunensine. (D) Sensograms comparing WT ACE2 and ACE2 expressed in the presence of 20 µM kifunensine followed by enzymatic digestion with endoH. (E) Sensograms comparing WT ACE2 with WT ACE2 enzymatically digested with fucosidase. (F) Calculated K_D for each of the ACE2 variants used, including repeats. Comparison of the % difference in K_D for each glycan modified ACE2 variant. Percentage difference was calculated for each glycan modified variant compared to a WT ACE2 sensogram performed during the same SPR experiment. The average of each repeat of this analysis is shown, with error bars representing +/- SEM shown where appropriate. ST6 ACE2, WT Sia ACE2, ST6 + ACE2 and Kif ACE2 were performed in triplicate whereas Kif + endoH and ACE2 + Fuc were performed once.

7.5 Perspectives

In this study, the impact of variable glycosylation on the interaction between ACE2 and SARS-CoV-2 S was investigated. By modifying the glycans using glycosyltransferases, glycosidases, and inhibitors of the glycan-processing pathway, it was possible to engineer glycan variants of soluble ACE2 and perform binding studies with trimeric SARS-CoV-2 S. Whilst the impact of fucosylation and the removal of the N-linked glycans was minimal, the removal and addition of sialic acids on ACE2 resulted in the increase and decrease of ACE2/SARS-CoV-2 S binding affinities, respectively. It was also interesting to note that kifunensine-treated ACE2 resulted in a decrease in binding affinity, which was recovered to the WT binding affinity when all but the first GlcNAc was removed by endoH. This indicates that there may be unfavourable interactions when the carbohydrate modifications on ACE2 are predominantly bulky $Man_9GlcNAc_2$ glycans.

These results suggest a limited role for the glycans of ACE2 in SARS-CoV-2 binding. When a glycan is critical for binding, the removal of the entire glycan will negatively impact binding [268]. Likewise when a glycan is shielding an antigenic protein surface the removal of the entire glycan will greatly increase the affinity to its binding partner [193]. In this case, the removal of glycans from ACE2 has minimal effects on SARS-CoV-2 binding, revealing that the glycans do not play a critical role in this interaction. What these studies suggest, however, is that the binding of ACE2 to SARS-CoV-2 S protein can be influenced by the particular composition of ACE2 glycans, albeit only to a modest degree. There is growing evidence for a role in sialic acid binding of SARS-CoV-2 on the cell surface, however this data suggests that the binding pocket for sialic acid is not productively utilized by ACE2 receptor glycans [122,383,384,394].

The use of recombinant ACE2 is being developed as a therapeutic against COVID-19 [378,395–398] and our results suggest that the SARS-CoV-2 S protein can bind to ACE2 promiscuously regardless of its glycosylation status. Whilst the study cannot address the impact of glycosylation *in vivo* or in a clinical setting this data suggests that therapeutic manufacture may be facilitated by a lack of strict glycan dependencies in its mode of recognition. However, it must be noted that sialylation of therapeutics can influence their half-life [121]. Overall, although ACE2 receptor glycosylation modestly impacts viral spike binding, it is highly likely that the glycosylation of virus and host target cells will have a significant impact upon the pathobiology of COVID-19 due to the extensive range of host lectins and the capacity of SARS-CoV-2 S protein to interact with host glycans. In conclusion, any role of glycosylation in the pathobiology of SARS-CoV-2 will lie beyond its immediate impact of receptor glycosylation on virus binding.

Chapter 8 Discussion, perspectives, and future work

8.1 The glycosylation of Env and S protein and how they differ

Using a combination of biophysical techniques this thesis aimed to provide in-depth determination of the glycosylation state of immunogens used to vaccinate against viral pathogens and compare and contrast the observed glycan signatures with those on their viral derived counterparts. Prior to the execution of this project little was known of how the glycosylation of recombinant immunogens compared to their viral derived counterparts at a site-specific level. Previous analyses included global analysis of Env produced in different cell lines, however, as broadly neutralizing antibodies bind to specific N-linked glycans it is important to compare the site-specific glycosylation of key bnAb epitopes, such as the N332 glycan supersite. During the course of the project several studies have additionally defined the glycan signatures of viral derived HIV-1 Env [163,264]. These studies supported the observations of the comparative site-specific glycan analysis of BG505 SOSIP.664 and BG505 viral-derived Env. The oligomannose-type glycans that dominate the glycosylation of the glycan shield are conserved across many regions of the viral spike, notably at the intrinsic mannose patch (IMP) centred around the N332 supersite and also at the trimer associated mannose patch (TAMP) which contains the key bnAb epitopes at N156 and N160. At sites flanking these regions, divergent processing was observed, with sites that presented a mixture of processing states on the SOSIP platform in both CHO and HEK293F cells such as N197 displaying elevated glycan processing on viral derived Env.

One key difference between viral derived and soluble recombinant SOSIP was the glycan occupancy at several sites across the protein. Closely spaced potential N-linked glycan sites (PNGS) on gp120 such as N133/ N137 and N190/N190c lacked glycan attachment on a significant proportion of sites on BG505 SOSIP.664 produced in HEK293F cells and in CHO cells. This is in contrast to the same sites on viral derived material which were fully occupied. Following this study, the underoccupancy of SOSIP glycosylation has been shown to influence the immune response elicited towards the immunogen. Macaque immunizations using BG505 SOSIP.664 generated neutralizing antibodies [186], however they were only able to neutralize BG505 and similar viruses. In addition to these analyses, epitopes for a number of non-neutralizing antibodies were defined at several regions on Env [299]. The most immunodominant non-neutralizing epitope was located on the base of the SOSIP trimer, which is exposed as a result of the truncation of the full length Env, required for the generation of soluble immunogens. However, an additional subset of non-neutralizing antibodies were isolated that bound to BG505 SOSIP.664, but were unable to

neutralize BG505 virus, unless the N611 glycan is knocked out. In addition to this, more recent studies have suggested that V1 glycan underoccupancy on SOSIP trimers may potentially induce non-neutralizing antibodies [399]. To close the artificial glycan holes and diminish the presentation of disadvantageous non-neutralizing antibody epitopes, the glycan sequons at regions that display underoccupancy on SOSIP trimers were modified to optimize their affinity to the glycosyltransferase enzymes. This approach demonstrated that it was possible to improve glycan occupancy across a number of different Env strains, without compromising the integrity of the glycan shield, and demonstrates that it is possible to modulate immunogen glycosylation towards more viral-like.

When comparing viral derived BG505 Env with BG505 SOSIP.664, produced in a research laboratory setting and under GMP for human clinical trials, subtle differences in the glycan shield were observed that have implications regarding the generation of non-neutralising antibodies. This highlighted the importance of applying the methodologies used to study HIV-1 Env to SARS-CoV-2 S protein. In a similar manner to vaccine design against HIV-1, many of the now authorized SARS-CoV-2 vaccines use a subunit vaccine-based approach, involving the presentation of the SARS-CoV-2 S protein prior to infection. As such it is important to establish the parameters that govern the presentation of N-linked glycans on the surface of the spike protein. With the determination of the three-dimensional structure of the SARS-CoV-2 S protein, the heterogeneous glycan shield was unresolved due to the flexibility of N-linked glycans. Using the same stabilized soluble 2P variant used to determine the structure of the S protein, the compositions of the glycans attached to the N-linked glycan sites were resolved. As with HIV-1 Env, regions of the glycan shield are sterically occluded from glycan processing and present elevated levels of oligomannose-type glycans not normally observed on healthy host glycoproteins. Despite this, the majority of N-linked glycan sites on S protein contained high levels of more processed, complex-type glycans, indicating that the glycan shield of SARS-CoV-2 S protein is less dense than HIV-1 Env. In addition, the glycan sites of S protein were highly occupied, in comparison to BG505 SOSIP.664.

To further compare the glycan shields of HIV-1 Env and SARS-CoV-2 S protein the glycosylation of soluble recombinant S protein from HEK293F and CHO cells were compared to the glycosylation of viral-derived S protein. This study demonstrated that, for SARS-CoV-2 S protein, the glycosylation of viral and recombinant S protein is consistent with respect to the site-specific presentation of oligomannose-type glycans, but that the processing of complex-type glycans differed. In addition, molecular dynamics simulations demonstrated a connection between the N-linked glycan processing state, with regards to the presentation of oligomannose-type glycans, and the accessible surface area of the glycan site to glycosidase enzymes. Correlating the accessible surface area to glycan

processing in this way demonstrates the differences in the glycan shields of HIV-1 and SARS-CoV-2. Whilst both S protein and Env possess oligomannose-type glycans, the abundance of oligomannose-type glycans was around half that observed on BG505 SOSIP.664 and demonstrates that the glycan shield of HIV-1 Env is denser compared to that of the SARS-CoV-2 S protein. This likely contributes to the relative ease with which patients infected with SARS-CoV-2 can elicit strong neutralizing antibodies targeting protein epitopes compared to HIV-1 infected individuals.

In addition to the SARS-CoV-2 S protein, the ACE2 receptor is also glycosylated. As such, the glycans on the ACE2 receptor must be considered when understanding the pathology of SARS-CoV-2. Using similar methodologies, the glycosylation of ACE2 was determined, displaying a diverse range of complex-type glycans at all sites analysed. By engineering the glycans of ACE2 using both enzymatic inhibitors and *in vitro* glycosidase treatment it is possible to investigate the role the glycans on ACE2 play in S protein binding. Differences in binding were observed when the sialylation of ACE2 was enhanced by ST6 co-transfection, however removal of the N-linked glycans did not ablate S protein binding, demonstrating that glycan microheterogeneity of ACE2 in the host cell will likely not provide a barrier to SARS-CoV-2 infection, specifically in the interaction between the S protein and ACE2. This work demonstrates that whilst glycosylation can be heterogeneous in places, there are certain trends and phenomena that are conserved across viral spike proteins, which can be informative for broader vaccine design strategies.

8.2 Observations of membrane bound vs recombinant HIV-1, composition, and occupancy

The glycosylation of viral envelopes differs from soluble immunogens on several fronts, each of which impacts the processing of the glycan shield, with hierarchies of consequences. The impact of membrane tethering on the glycosylation of HIV-1 Env is seen across the trimer, such as the higher abundance of sialic acid containing glycans which indicates that membrane tethering brings Env in closer proximity to the glycan processing enzymes. This effect is also contributing to the increased activity of GlcNAc transferases resulting in increased branching at sites populated by complex-type glycans and membrane tethering also likely results in enhanced oligomannose-type glycan processing of sites presenting a mixture of both complex-type glycans and oligomannose-type glycans, such as N197, on soluble forms of Env. These phenomena were also observed in a study comparing recombinant SOSIP with recombinant full length (FL) Env, and UPLC analyses revealed that the complex-type glycans were larger, more branched and more sialylated on the full length material compared to the SOSIP [301]. In addition, site-specific analysis revealed that mixed sites such as N197 were processed analogously to viral derived Env,

even though the Env used in the cited study is the clade B isolate AMC011 and the virus was BG505. Other studies have observed larger branched glycans on membrane tethered Env, confirming this observation [163,264]. Interestingly, whilst the SARS-CoV-2 viral derived S protein did display elevated processing at sites such as N149 and N165 compared to the averaged recombinant S protein data (**Figure 6-1**), the elevation in glycan processing was not as extensive as observed when viral derived Env glycosylation was compared to that of SOSIP. Unlike viral-derived Env, S protein produced from infectious virions did not display enhanced glycan branching and possessed almost no sialylation at every site analysed (**Figure 6-2** and **Appendix G.2**). One potential explanation for this is that the egress from the ER and Golgi apparatus differs for SARS-CoV-2 and HIV-1. It has been established that the SARS-CoV-2 virus buds into the ERGIC, the presence of Golgi-specific glycan modifications such as galactosylation implies that the virions do not exit the secretion pathway at this stage, but rather buds into the ERGIC lumen [400]. This will effectively ablate the effect of membrane tethering keeping the nascent glycan sites in close proximity to the glycan processing enzymes, and therefore reduces both the branching and sialylation of the viral derived S protein (**Figure 8-1**).

It is important to note that other reasons for these differences could be contributing to the different glycan processing states between the two viruses. The producer cell for the SARS-CoV-2 and HIV-1 viruses used in this thesis were different, and therefore the expression levels of glycosyltransferases may be different, which could impact the glycan processing. The AMC011 FL construct described above was produced in HEK293 cells alongside the SOSIP and the correlation with this analysis and the BG505 analysis likely means that the cell line is not the main driver of the changes observed between membrane tethered Env and soluble SOSIP. The other difference that could be contributing to the observed changes in glycosylation is that the stabilising mutations introduced into both the SOSIP and 2P soluble glycoproteins are stabilising the soluble structures beyond those on their viral counterparts, effectively generating a more stable quaternary structure that reduces the accessible surface area of the PNGS of the soluble variants. This is difficult to control for, as removal of the stabilising mutations creates a metastable protein that is no longer native-like. However, the oligomannose-type glycan processing of SARS-CoV-2 is analogous for both soluble and viral S protein, suggesting that the introduction of the 2P mutations is not affecting glycan processing. Further work will be needed to address the impact of stabilizing mutations on glycan processing, including broadening the number of viral comparisons. The trends observed for full length Env have only been determined on a handful of viruses, and a full understanding of the glycosylation of viruses will need to investigate the glycosylation of infected patient-derived virus, although the yield of virus would be too small to obtain high quality data using the methods outlined above.

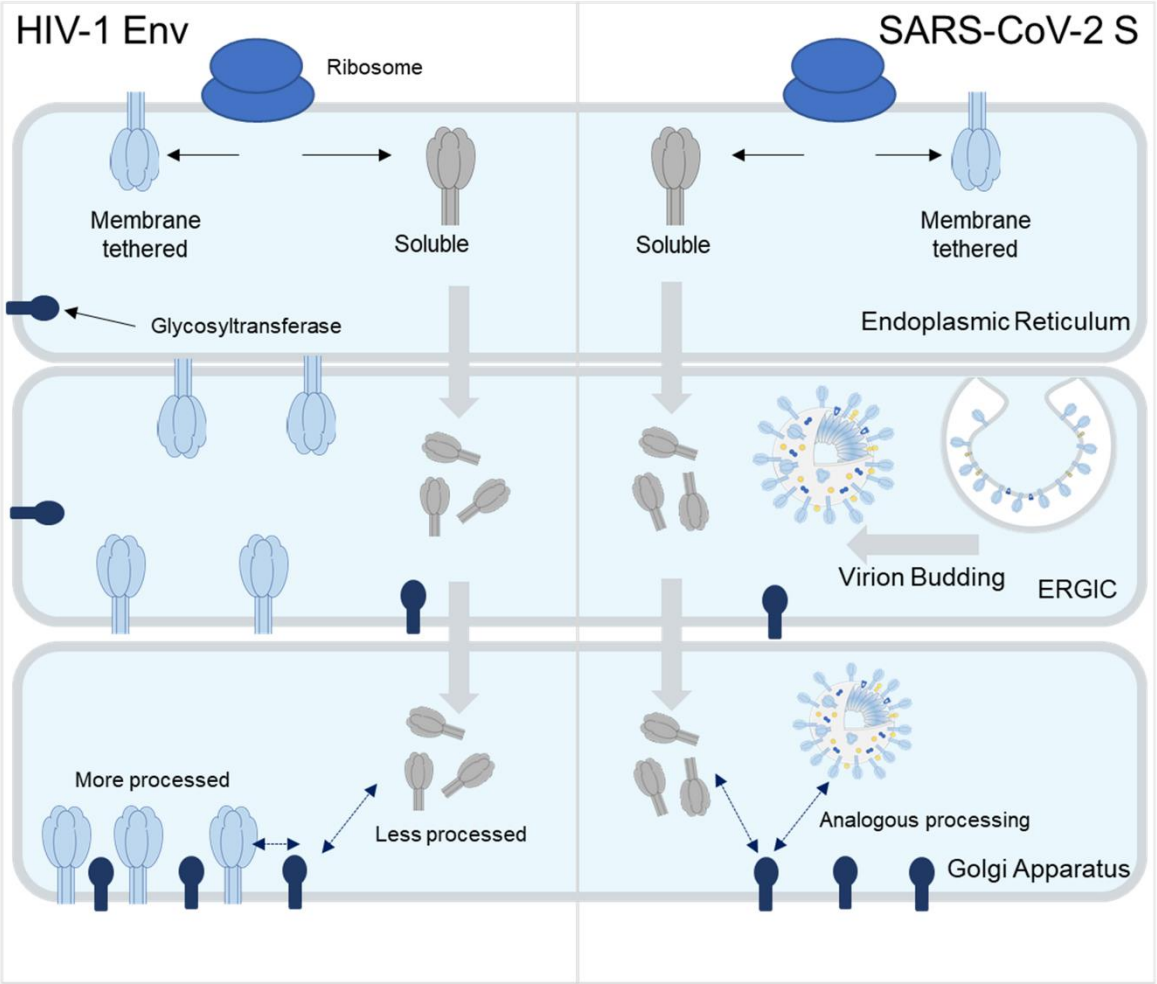


Figure 8-1:Rationalizing the differences in S protein and Env glycosylation resulting from the presence/ absence of the transmembrane portion of the protein.
Representative cartoon displaying the differences in how soluble, and membrane tethered virions interact with glycosyltransferases. Whereas HIV-1 Env assembles on the plasma membrane, SARS-CoV-2 virions bud into the ERGIC, this effectively distances the S protein from the glycosyltransferase enzymes, reducing the complex-type glycan elaboration that can occur, such as sialylation, to a similar extent to soluble Env and S protein.

8.3 Repairing holes in the glycan shield is an important consideration in vaccine design

In addition to localized differences in glycan processing, differences in site-specific glycan occupancy were observed between recombinant and viral envelope glycoproteins. Throughout the course of this project studies on HIV-1 Env immunizations using SOSIP trimers has revealed the difficulties created by artificial glycan holes in that they result in the generation of epitopes on the soluble immunogen that are not present on the virus, and therefore result in the elicitation of non-neutralizing epitopes [218,299]. Glycan holes can arise on HIV-1 Env through a number of routes; firstly, a glycan hole that is strain specific due to a lack of a particular N-linked glycan that is abundant in the circulating viral population, such as the N241/ N289 glycan hole present on BG505 [281]. In addition to strain-specific glycan

holes, this work demonstrated that glycan sites on BG505 are differentially occupied between SOSIP and virus. These were termed artificial glycan holes due to their absence on infectious virus. In addition, recombinant SARS-CoV-2 S protein displayed regions of under occupancy, however these were much less apparent than for HIV-1 Env. These studies showed that glycan holes can arise for many reasons, but that it is possible to enhance glycan occupancy of soluble trimers by modifying the N-linked glycan sequon for particular sites. Unoccupied PNGS appear as a result of a combination of three features. Firstly, closely spaced glycan sites, such as N190 and N190c, display underoccupancy, likely due to the difficulty for the glycosyltransferase enzymes that add the precursor N-linked glycan from the dolichol phosphate to the NxS/T to both sites before the nascent polypeptide chain moves away from the ribosome/ER translocon complex as a result of continued translation. This phenomenon is not observed on SARS-CoV-2 S protein as there are no sites which are this closely spaced, the two closest are 8 amino acids apart at N709 and N717.

This alone cannot explain the reduced occupancy of SOSIP trimers as the viral counterpart has high occupancy at both N190 and N190c and therefore other factors play a role as well. The third position of the N-linked glycan sequon impacts glycan occupancy, as can be seen at several sites, most noticeably N611, whereby the NxS N-linked glycan sequon has reduced affinity to the glycosyltransferase machinery and glycan attachment is less likely. Converting the NxS glycan sequons to NxT generally increased N-linked glycan occupancy, however if this dictated glycan occupancy alone then likely the immune selection pressure would mean that viral Env would contain exclusively NxT glycan sequons to ensure maximum occupancy. In addition, SARS-CoV-2 S protein contains several NxS sites, and these are highly occupied on both viral and soluble material.

Finally, artificial glycan holes are more prominent towards the C-terminus of the viral spike. On HIV-1 Env gp41 displays decreased glycan occupancy compared to the viral counterpart at all four sites, depending on the strain. In **Chapter 6** recombinant soluble S protein from several labs was compared. The Amsterdam, Harvard, Swiss, and to a lesser extent Oxford S protein possessed moderately reduced occupancy at sites proximal to the C-terminus of the protein, specifically N1173 and N1198, or N1134 for the Amsterdam S protein, as this soluble construct was truncated prior to the last three glycan sites. For both SARS-CoV-2 and HIV-1 viral derived glycoproteins the same C-terminal sites were fully occupied. Whilst N611 was not resolved in the HIV-1 Env virion analysis the ability of N611-glycan hole binding Abs to neutralize BG505 virus lacking the N611 site provides evidence that this site is highly occupied. These artificial glycan holes are likely arising due to the truncation of the C-terminus required to remove the transmembrane region. As the nascent polypeptide finishes translation it subsequently detaches from the translocon complex and

therefore the glycosyltransferases have reduced capacity to attach N-linked glycans to sites closer to the C-terminus [362]. The presence of an extended C-terminus on viral derived glycoproteins likely enhances glycan occupancy of these C-terminal N-linked glycan sites compared to the solubilized equivalents.

In addition to direct influences from the amino acid composition, broader parameters contribute to the differences in occupancy between viral and soluble glycoproteins. In order to produce large amounts of glycoproteins required for biophysical assays and clinical trial, a process known as codon optimization is used. As the same amino acid can be coded for by different RNA sequons, there are multiple tRNA within the cell that have the same amino acid attached with different RNA sequences. The abundance of tRNAs vary between organisms, and it is possible to optimize the codon usage so that the most abundant tRNAs for a particular amino acid are used during translation. The more abundant tRNAs are more likely to bind to a ribosome during translational elongation and therefore the elongation process progresses faster the more codon optimized a gene is [401]. As the attachment of N-linked glycans occurs co-translationally, the longer the process of elongation takes the longer the glycosyltransferase enzymes have to attach an N-linked glycan at a PNGS. As codon optimization is required for the production of high yields of recombinant material, this could contribute to the reduced occupancy of soluble recombinant envelope glycoproteins compared to their viral counterparts. In addition the *env* gene in infectious virions has been shown to possess a bias towards rare codons, which will slow the elongation process, further enabling high levels of glycan occupancy to be achieved [401].

These analyses have revealed the importance of monitoring candidate immunogens for artificial glycan holes. To fully map the elicited antibody response of HIV-1 Env it is essential to define any glycan holes, especially considering the modern germline targeting approaches involves glycan deletions. An additional example of a germline targeting approach developed is fusion peptide bnAb germline targeting, to attempt to elicit PGT151-like bnAbs. Fusion peptide targeting antibodies are able to recognize and neutralize virus lacking the N611 glycan site [402,403], and germline targeting approaches will likely involve priming the immune system with an immunogen lacking N611. In a subsequent boost immunogen, the N611 site will need to be reintroduced, as the majority of Env strains contain this site. If this booster immunogen is under occupied at N611 then the antibody response will not evolve to accommodate the N611 glycan and the germline targeting approach will not be successful, highlighting the need to monitor and repair artificial glycan holes in the glycan shield. Further efforts are required to silence the glycan hole located at the base of the SOSIP trimer and the introduction of N-linked glycans on the base of the trimer is one approach to achieve this. In addition, it will be important to identify any glycan holes in material produced by the host following administration of nucleotide-based

vaccines, such as mRNA, for both HIV-1 Env and SARS-CoV-2, to probe the integrity of the glycan shield and ensure the best chance of success for potential vaccine candidates.

8.4 Atypical glycan processing is a conserved feature of viral spike glycoproteins.

The presence of elevated levels of oligomannose-type glycans on SARS-CoV-2 and HIV-1 envelope glycoproteins reveal a shared feature of viral glycan shields that is distinct from typical host glycoprotein. The work outlined in **Chapter 6** demonstrates the steric protection that regions of the SARS-CoV-2 glycan shield are under, using a combination of site-specific glycan analysis and molecular dynamics modelling. The correlation between the accessibility of a glycan site and it's processing state has been shown for HIV-1 Env, with poorly accessible sites correlating with those that present high levels of oligomannose-type glycans [153,404]. Additional studies have highlighted this phenomenon for other viral glycoproteins, notably Influenza Haemagglutinin (HA) from H3N2 and Lassa virus GPC [323,405]. As the glycan shield becomes denser, the processing of N-linked glycans decreases, resulting in a weakness in the glycan shield that provide distinction between host cell glycan processing. Additionally, the evolution of the H3N2 glycan shield has revealed a selective pressure to introduce N-linked glycan sites onto the immunodominant head region of HA, which has increased during the endemic circulation of the virus between 1968 and 2011. This resulted in a concomitant increase in under processed oligomannose-type glycans on HA [405]. In addition, HIV-1 displays similar trends, albeit over a much shorter timescale due to the high mutation rate of the HIV-1 genome. Within a single infection cycle, the number of PNGS on Env can vary, with initial transmitter/founder strains possessing fewer N-linked glycans, followed by a steady increase as the infection becomes chronic, and finally a decrease as the immune system fails [145]. The glycan shield was likely present prior to species crossover to humans. A recent study investigating the glycosylation of a SOSIP trimer of a simian immunodeficiency virus (SIV) termed MT145k [406] demonstrated remarkable conservation in the glycan shield between HIV and SIV Env, with respect to the presentation of oligomannose-type glycans. The glycan shield, therefore, likely plays an important role in avoiding the host immune response in multiple organisms.

These observations highlight the shielding role that N-linked glycans perform when evading the host immune system. However, for SARS-CoV-2, the position and number of N-linked glycan sites has, so far, remained consistent, with mutations more common around the RBD [407]. One potential reason why glycan site additions are less prevalent is the potential fitness cost to viral infectivity resulting from glycan addition. A recent study demonstrated that genetically similar coronaviruses contain a glycan at position N370 and

that the loss of the N370 glycosylation site in SARS-CoV-2 may have increased the availability of the “open” S form, which may contribute to its higher infectivity relative to SARS-CoV1 and other variants containing the N370 site.

In addition to potential fitness costs of additional glycan sites to the virus, disease mechanisms may explain why SARS-CoV-2 does not have an extensive glycan shield. Unlike HIV-1, SARS-CoV-2 is an acute infection, whereby infectivity is at its peak over a relatively short period. This means that viruses are able to transfer to a new host quickly without a protracted interaction with the host immune response, and the remaining viruses are cleared by the immune system. As such SARS-CoV-2 S protein does not have the same selection pressures as HIV-1 Env, which can exist for years within the host. Over time, however, as the immune memory of the population expands, there is the possibility that the virus will evolve a denser glycan shield. Coronaviruses have been circulating in the human population far longer than the outbreaks in the 21st century. A portion of the common cold viruses are coronaviruses, including NL63 and 229E [408,409]. Interestingly the spike proteins of HCoV-229E and HCoV-NL63 are of a similar size to SARS-CoV-2 S protein, however they contain a significantly larger number of N-linked glycan sites compared to SARS-CoV-2 with HCoV-NL63 containing 39 and HCoV-229E possessing 30 per protomer, compared to 22 on the SARS-CoV-2 S protein. These viruses have been circulating in the human population for longer than SARS-CoV-2 [410]. As with H3N2, it appears that circulation within the human population has resulted in the addition of N-linked glycan sites to evade the immune system. As SARS-CoV-2 continues to circulate and mutate, the chances of introducing a potential N-linked glycan site increase and demonstrate the key role that the glycan shield plays in shielding the antigenic protein surface. This work demonstrates that even relative sparse glycan shields, such as that of SARS-CoV-2, display regions of restricted glycan processing, such as at N234. This demonstrates a conserved feature of viral glycan shields that may be exploited by anti-mannose lectins in potential future viral therapeutics [405,411].

8.5 The differential importance of N-linked glycosylation in vaccine design

Whilst over time the glycan shield of viruses may evolve and increase in density, what matters most for vaccine design is protecting against immediate circulating strains and closely related mutants. The recent outbreak of SARS-CoV-2 and the rapid development and success of SARS-CoV-2 vaccines, as well as the strong patient immune response, demonstrate a differential importance of N-linked glycosylation in vaccine design compared to HIV-1 Env. Whilst a knowledge of the glycan shield of SARS-CoV-2 is useful as it demonstrates potential glycan clusters that may shield particular regions of the antigenic

spike protein, recent vaccines have been developed which clearly demonstrate that the particular composition of N-linked glycans on the S protein will have a limited role in governing the antibody-mediated immunity to SARS-CoV-2. This is best exemplified in the apparent success of the Novavax vaccine candidate, NVX-CoV2373, which demonstrated an efficacy of ~85% for the B.1.17 (Kent/alpha) SARS-CoV-2 variant [412]. This vaccine is produced in moth cells and will therefore be glycosylated and processed by an insect N-linked glycan processing pathway [350]. The early stages of oligomannose-type glycan processing are conserved between insects and mammalians, however complex-type N-linked glycan processing, such as the addition of sialic acids, is not conserved and the resultant compositions of sites presenting high abundances of complex-type glycans are distinct from protein produced in mammalian cells [349]. Despite this the vaccine is able to provide protection with an analogous efficacy as other vaccine immunogens which are produced by human/mammalian cells and demonstrate that the glycan shield of SARS-CoV-2 is not a barrier to generating an immune response.

Despite the glycan shield not providing a barrier for the induction of effective antibody-based immunity the N-linked glycans of SARS-CoV-2 S protein are still important for other reasons. Molecular dynamics simulations and binding assays have explored the role of the N234 and N165 glycans stabilizing RBD conformations [346]. Glycans are also critical for stabilising the structure of the SARS-CoV-2 S protein, with a recent study reporting that N801Q and N1194Q S protein mutants resulted in severely decreased protein yield, including protein degradation [413]. Additionally, the role of glycosylation in SARS-CoV-2 infection is still under investigation. This thesis demonstrated that modifying the glycans on the ACE2 receptor can modulate SARS-CoV-2 S protein binding, however they are not critical for binding as the near complete removal of the N-linked glycan did not ablate ACE2 binding. Additional studies have shown that modification of SARS-CoV-2 glycans using similar approaches can impact ACE2 binding [413]. Taken together, the N-linked glycans on the SARS-CoV-2 S protein are required for correct protein folding and function, however, in vaccine design the glycan shield does not act as a barrier to providing antibody-mediated immunity as the glycan shield is not dense enough to obscure the underlying antigenic protein surface, as can be seen in the conserved low levels of oligomannose-type glycans on S protein, relative to other viral spike glycoproteins.

Unlike SARS-CoV-2 S protein the N-linked glycans of HIV-1 Env play a critical role in the induction of antibody-mediated protection against infection. The N-linked glycans comprise half the mass of Env and can contain over 30 glycans per protomer. In addition, the Env ectodomain consists of ~700 amino acids, whereas S protein consists of over 1200 and as such the density of the HIV-1 Env far exceeds that of S protein. This can be seen in the high levels of oligomannose-type glycans across multiple regions of the protein, with

many sites lacking any further processing beyond the initial stages of N-linked glycan maturation. The critical role of the dense glycan shield in evading immune surveillance can be seen in its cross-species conservation between SIV and HIV [406]. Additionally when holes appear in the glycan shield, a strong immune response is capable of targeting the region underneath the glycan, with these epitopes forming the dominant targets of the antibody response following immunization [281]. Due to its importance in shielding the more antigenic protein surface, the glycan shield is one of the more conserved features of Env. Sites such as N332 are present on the majority of Env strains, and for every strain analysed in this study this site contains exclusively oligomannose-type glycans (**Figure 4-5**). A combination of this cross-clade conservation, and the distinct glycan processing from host glycoproteins, means that, in a subset of infected patients, broadly neutralizing antibodies have been elicited, with almost all of them either directly recognizing a glycan or accommodating the presence of a glycan as part of their epitope [414]. Although the majority of glycan binding antibodies recognize the mannose clusters in the IMP and TAMP, other antibodies such as PGT151 display preferences for complex-type glycans, specifically tri- and tetra-antennary glycans [107]. Whilst not all classes of bnAbs will be elicited, the glycan compositions and locations will dictate the success of many vaccination approaches against HIV-1. The work outlined in **Chapter 4** demonstrated the role that artificial glycan holes can have in distracting the immune system from potential neutralizing responses and demonstrate the importance of the work outlined in this study in the vaccine design process against HIV-1.

8.6 Future work

This thesis aimed to explore the differences between the glycan shield of SARS-CoV-2 and HIV-1 envelope glycoproteins, and how they compare to their viral counterparts. Due to sample availability and requirements viruses were produced in cell lines. As the glycosyltransferase expression levels can differ between cell types the next stages of this work will include analysis of virus from an infected individual. In addition to different glycosyltransferase expression rates there may be additional stresses on cells in a patient that are not present in cells in culture, for example extracellular signalling from the immune system, that could be influencing the processing of complex-type glycans in the glycan shield. This thesis has demonstrated that the presentation of oligomannose-type glycans on SARS-CoV-2 S protein and HIV-1 Env are driven by the steric protection afforded by either neighbouring glycans, or through protein clashes, and that these features will likely be conserved, despite the differences outlined above.

To date only a handful of viral-derived Env proteins have been analysed using site-specific analysis. Broadening the number of clades and strains with which the site-specific glycosylation is known will further inform vaccine design. An example of key strains of

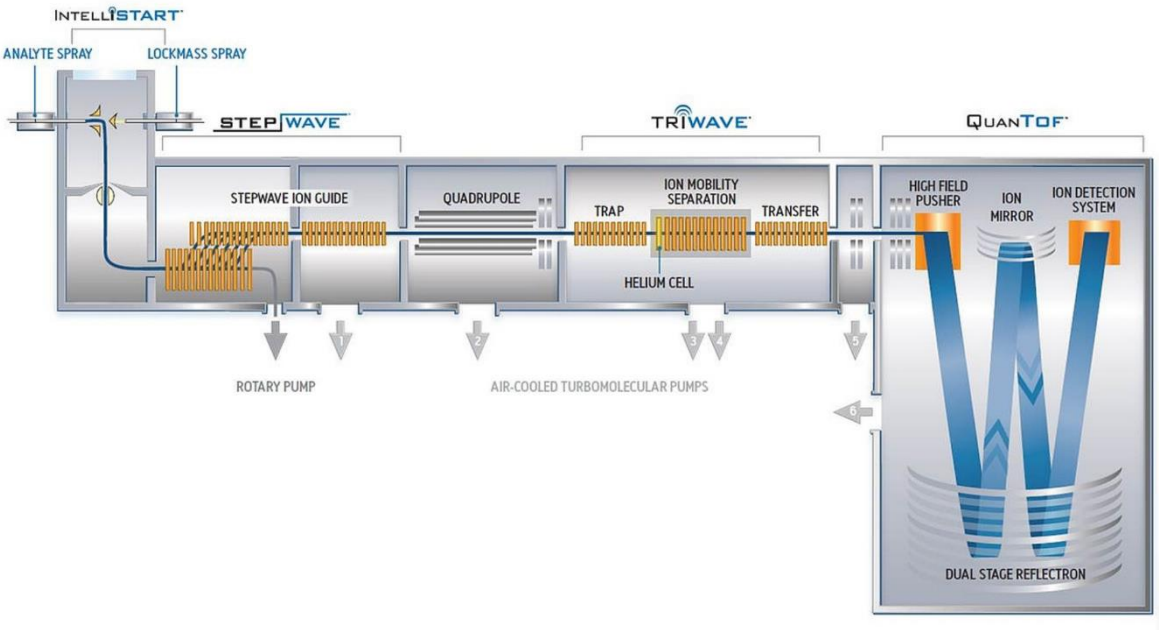
interest includes those viruses which are resistant to bnAb neutralization. A recent clinical trial investigated the ability of a single monoclonal antibody, VRC01, to prevent HIV-1 infection [350]. Whilst some efficacy was observed, there were several breakthrough infections and work is currently ongoing to identify what features of the virus enable the evasion of broad and potent bnAbs, and the glycan shield will likely be playing a role in this resistance. Furthermore, as more subunit vaccines enter clinical trials it will be important to monitor both the glycan composition and the presence of artificial glycan holes to rationalize the antibody response elicited towards the immunogen.

Finally, the vaccine development of SARS-CoV-2 has demonstrated that new vaccine technologies offer the potential for the rapid design and testing of new vaccines. A key innovation is the use of nucleotide-based vaccines whereby the vaccine recipient produces the immunogen within their own cells. As a result, novel approaches are required to define the glycosylation of the material made by nucleotide-based vaccines to ensure that they are consistently glycosylated to material produced recombinantly. This will range from the *in vitro* production of viral immunogens into particular cell lines relevant to the sites of vaccination, such as muscle cells. Beyond this, as the mass spectrometric methods become increasingly sensitive, it will be important to investigate the post-translational processing of immunogen extracted from lymph nodes of vaccine recipients to investigate the exact glycan compositions presented to the recipient's immune system, as these compositions will be the ones that are driving the evolution of vaccine induced immunity. The analysis of viral envelope glycoproteins is integral to understanding the antigenic surface of the protein. Even though the precise processing and presentation of N-linked glycans requires less attention for certain viruses, such as SARS-CoV-2, it is impossible to know ahead of time whether or not the glycan shield will provide a barrier to vaccine design strategies, and as such the development of a broadly applicable workflow to define the glycosylation of viral spike proteins that can be rapidly deployed to newly emergent viruses remain crucial to modern vaccine design strategies.

Appendix A List of non-structural proteins in SARS-CoV-2 and their molecular functions.

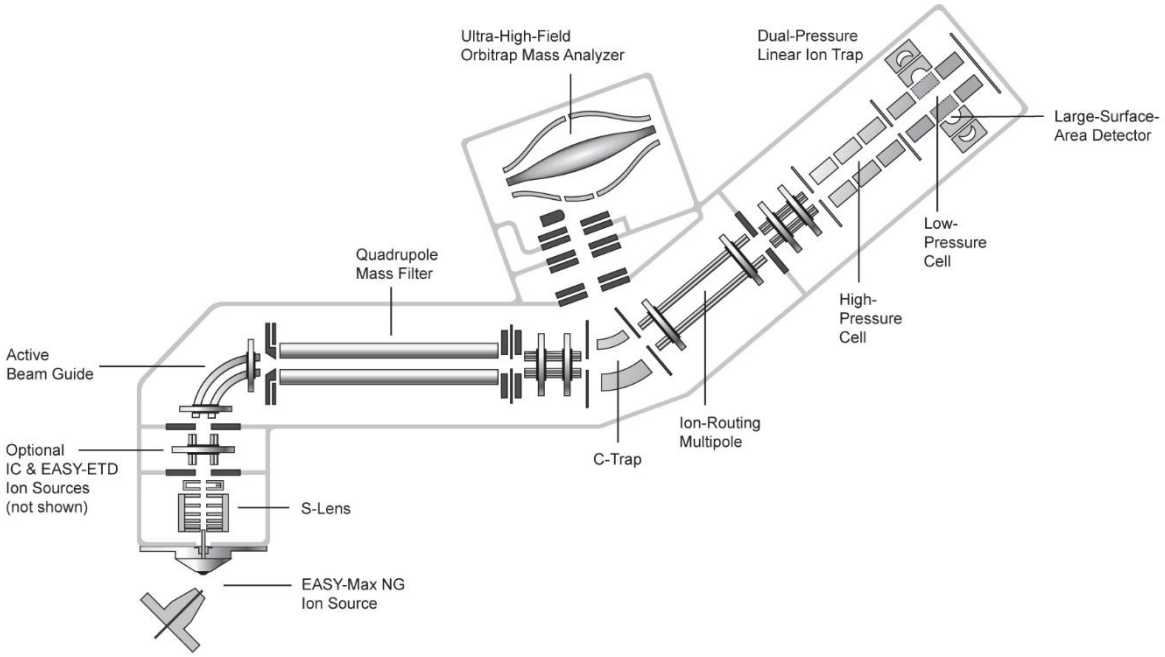
Protein name	Description	Proposed function
Nsp1	Nsp1 is the N-terminal product of the viral replicase	Leader protein host translation inhibitor. Mediates RNA replication and processing. Involved in mRNA degradation.
Nsp2	Nsp2 is a replicase product essential for proofreading viral replication	Modulation of host cell survival signalling pathway by interacting with host PHB and PHB2.
Nsp3	Nsp3 is a papain-like proteinase contains several domains.	Functions as a protease to separate the translated polyprotein into its distinct proteins.
Nsp4	A membrane-spanning protein contains transmembrane domain 2 (TM2)	Believed to anchor the viral replication-transcription complex to modified ER membranes.
Nsp5	3C-like proteinase and main proteinase	Involved in viral polyprotein processing during replication.
Nsp6	Putative transmembrane domain	Plays a role in the initial induction of autophagosomes from host endoplasmic reticulum.
Nsp7	Nsp7 is an RNA-dependent RNA polymerase	It forms a hexadecameric super-complex with nsp8 that adopts a hollow cylinder-like structure implicated replication.
Nsp8	Multimeric RNA polymerase; replicase	It forms a hexadeca-meric super-complex with nsp7 that adopts a hollow cylinder-like structure implicated replication.
Nsp9	A single-stranded RNA-binding viral protein	Participate in viral replication by acting as an ssRNA-binding protein.
Nsp10	Growth-factor-like protein contains two zinc-binding motifs	In viral transcription by stimulating both nsp14 3′-5′ exoribonuclease and nsp16 2′-O-methyltransferase activities. Plays an essential role in viral mRNAs cap methylation.
Nsp12	RNA-dependent RNA polymerase (Pol/RdRp)	Responsible for replication and transcription of the viral RNA genome.
Nsp13	Zinc-binding domain, NTPase/helicase domain, RNA 5′-triphosphatase	A helicase core domain that binds ATP. Zinc-binding domain is involved in replication and transcription.
Nsp14	Proofreading Exoribonuclease domain (ExoN/nsp14)	Exoribonuclease activity acting in a 3′ to 5′ direction and N7-guanine methyltransferase activity.
Nsp15	EndoRNAse; nsp15-A1 and nsp15B-NendoU	Mn(2+)-dependent endoribonuclease activity
Nsp16	2′-O-ribose methyltransferase	Methyltransferase that mediates mRNA cap 2′-O-ribose methylation to the 5′-cap structure of viral mRNAs.
Nsp11	Made of 13 amino acids (sadaqsflngfav) and identical to the first segment of Nsp12.	Unknown

Appendix B Schematic of Waters Synapt G2Si used for released glycan composition and structural analysis



Schematic obtained from Waters Synapt G2Si Brochure [415]

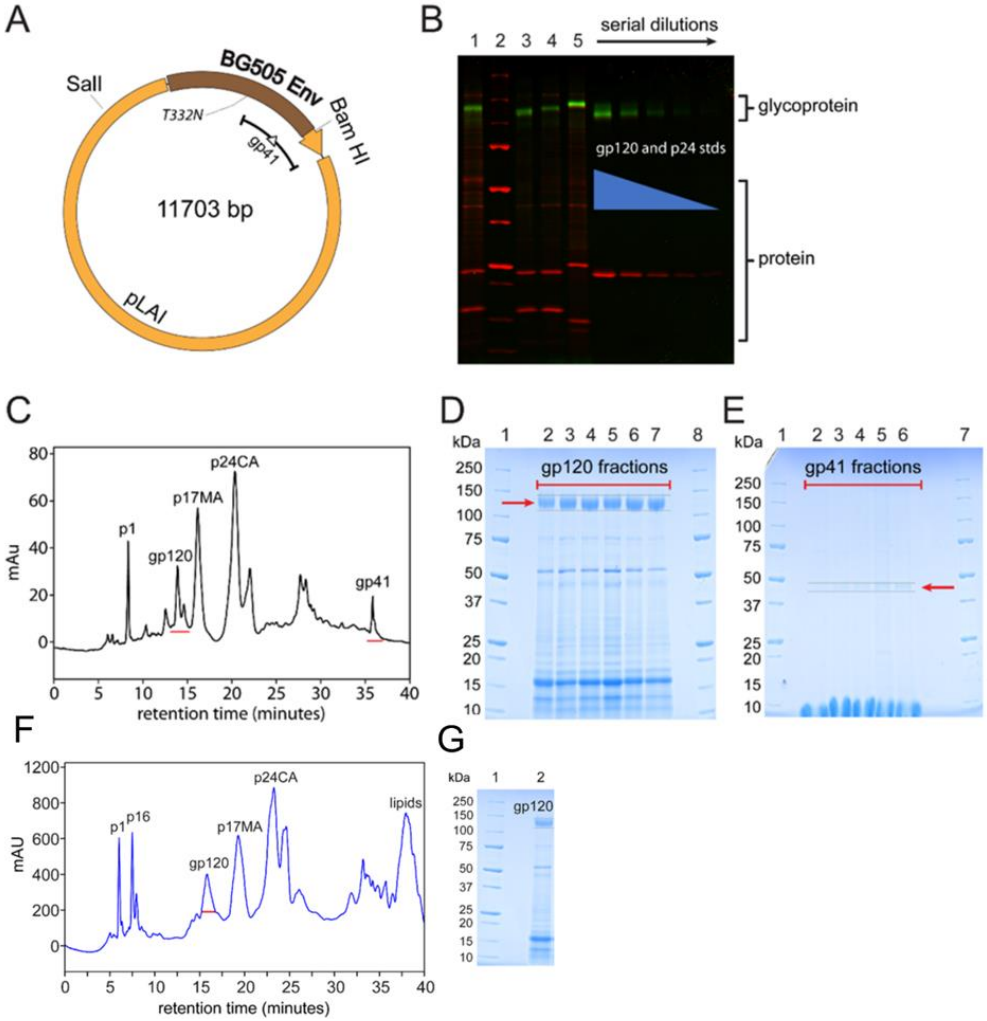
Appendix C Schematic of an Orbitrap Fusion mass spectrometer used for site-specific glycan analysis



Schematic obtained from "Planet Orbitrap", a web resource detailing the applications and uses of orbitrap mass spectrometers [416].

Appendix D Supplemental information for Site-Specific Glycosylation of Virion-Derived HIV-1 Env Is Mimicked by a Soluble Trimeric Immunogen.

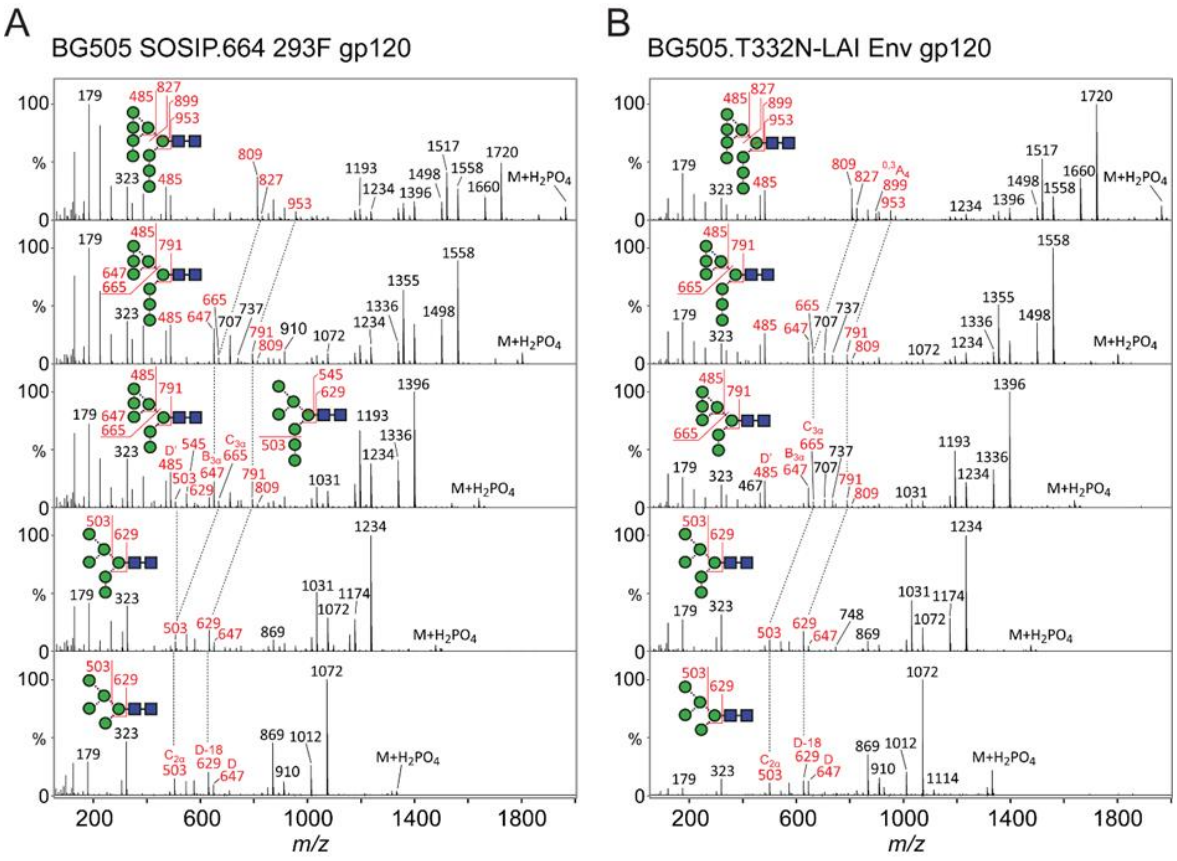
D.1 Construction and purification of the BG505.T332N-LAI infectious molecular clone (IMC) and its Env proteins.



(A) Schematic representation of the BG505.T332N-LAI IMC. (B) Lane 1: Two-dye SYPRO-stained (total protein, red; glycoproteins, green) SDS-PAGE gel analysis of BG505.T332N-LAI. Lane 2: M.W. standards. Lanes 3–5: Positive control samples (3: HIV-1 BAL, 4: HIV-1 NL4/3, 5: SIVmac239). The gel-staining was calibrated using the gp120 (200–12.5 ng) and purified p24 (600–37.5 ng) standards shown on the right. (C) HPLC separation of BG505.T332N-LAI/A66-R5 (P4408) under non-reducing conditions. UV absorbance was measured at 206 nm. Viral protein peaks (identified by subsequent SDS-PAGE gel, protein sequencing and immunoblot analyses) are labelled above the chromatogram. (D) SDS-PAGE gel of gp120. Lanes 1 and 8: M.W. standards, Lane 2-7: pooled HPLC fractions underscored in red on panel C. (E) SDS-PAGE gel of gp41. Lanes 1 and 7: M.W. standards, Lanes 2-6: pooled HPLC fractions are underscored in red on panel C. The gel bands in D (gp120) and E (gp41) indicated with red arrows were excised and used for released glycan analysis by UPLC and mass spectrometry. (F) HPLC separation of BG505.T332N-LAI/A66-R5 (P4408) under non-reducing conditions for site-specific analysis. Viral protein peaks (identified by subsequent SDS-PAGE gel, protein sequencing and immunoblot analyses) are labelled above the chromatograph. (G) SDS-PAGE gel of gp120. Lane 1: M.W. standards, lane 2: pooled HPLC fractions underscored in red on panel A. This HPLC-purified material was subsequently used for site-specific glycan analysis. External

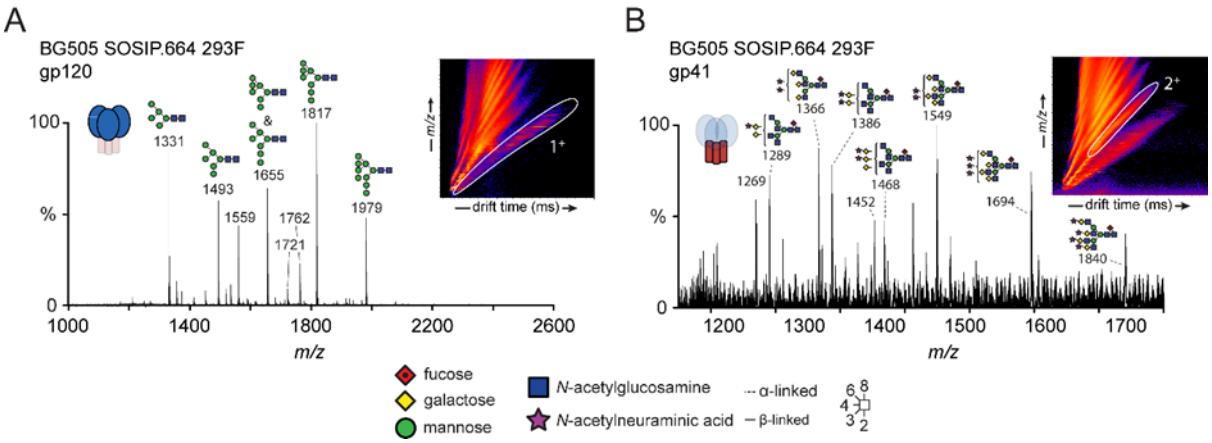
collaborators from the laboratory of Prof. Lifson (Frederick National Laboratory, USA) assisted this study by producing and purifying the viral Env

D.2 MS/MS fragmentation spectra of oligomannose-type glycans to determine their fine structure.



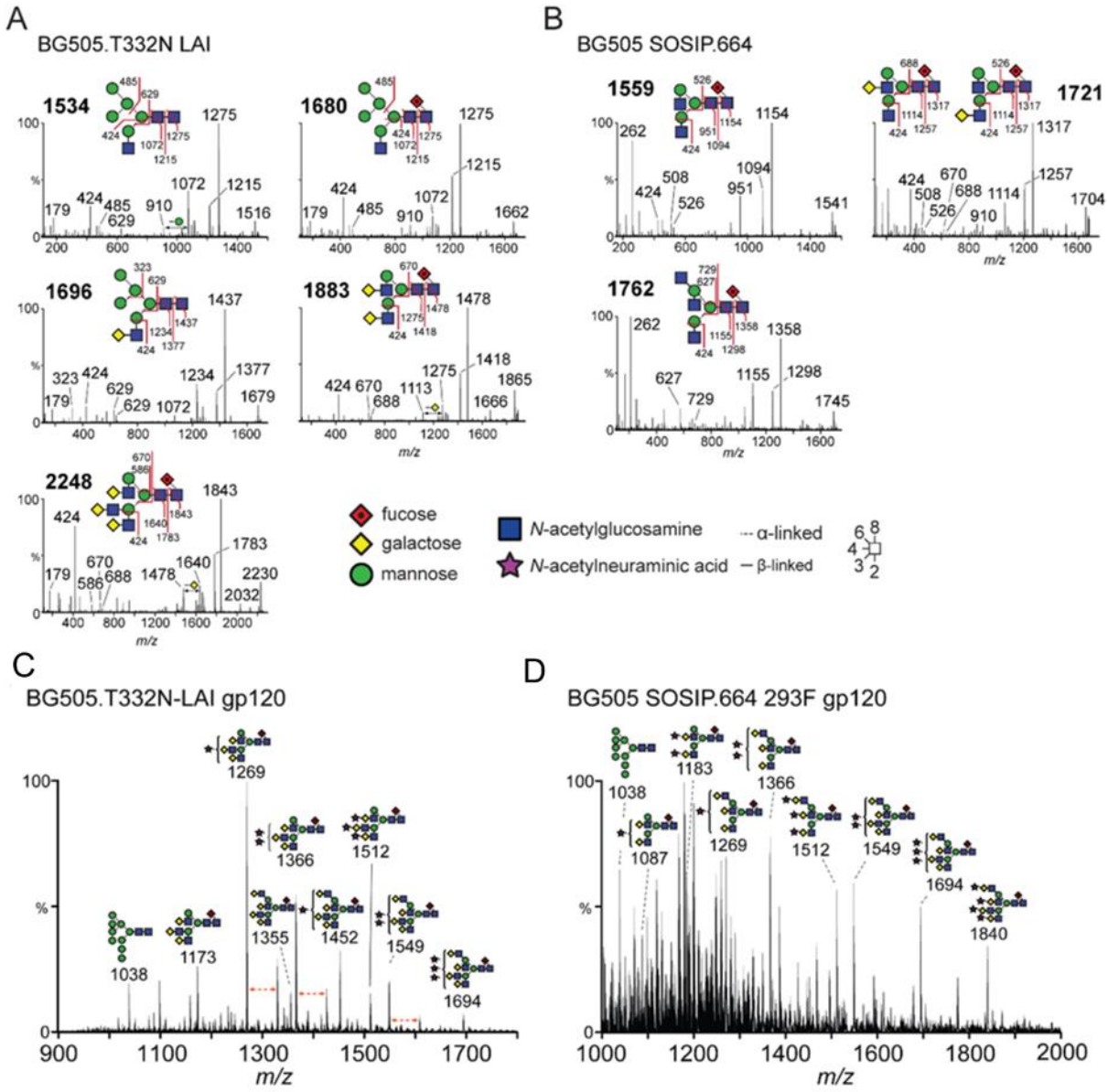
(A) Fragmentation spectra of oligomannose-type glycans released from 293F-derived BG505 SOSIP.664 with diagnostic structural ions highlighted in orange. (B) Fragmentation spectra of oligomannose-type glycans released from 293F-derived BG505.T332N-LAI with diagnostic structural ions highlighted in orange. Peaks are annotated according to the scheme proposed by Domon and Costello [127].

D.3 Ion mobility mass spectrometric analysis of gp120 and gp41 N-glycans released from BG505 SOSIP.664 produced in 293F cells.



(A) Mobility-extracted singly charged negative ions found on BG505 SOSIP.664 gp120. The corresponding singly charged ions are encircled with the white oval in the DriftScope image (*m/z* against drift time). (B) Mobility-extracted doubly charged negative ions found on 293F-derived gp41. The corresponding doubly charged ions are encircled with the white oval in the DriftScope image (*m/z* against drift time).

D.4 MS/MS fragmentation spectra of BG505.T332N-LAI gp120 and BG505 SOSIP.664 complex N-glycans.



(A) MSMS fragmentation spectra to determine the fine structure of the top 5 most common singly charged ions corresponding to complex-type glycans released from BG505.T332N-LAI gp120. (B) MSMS fragmentation spectra to determine the fine structure of the top 3 most common singly charged ions corresponding to complex-type glycans released from BG505 SOSIP.664 gp120 produced in 293F cells. (C) Doubly charged N-glycans extracted in Driftscope based on their *m/z* and retention time through the ion mobility cell from gp120 derived from BG505.T332N-LAI virions. Peaks with additional phosphate adducts are labelled in orange. (D) Doubly charged N-glycans from gp120 derived from BG505 SOSIP.664 produced in 293F cells.

D.5 Calculated Gag:Env ratios for the HIV-1 LAI BG505.T332N/A66-R5 P4408

sample	gag:env	~trimer
P4408 HIV-1 LAI BG505.T332N/A65-R5	24	20
P4239 HIV-1 BAL/SupT1-R5	16	28
P4249 HIV-1 NL43/SupT1	42	11
P4004 SIVmac 239/SupT1-R5	10	48

Preparation and reference standard virus preparations based on densitometric measurements, normalized for molecular weights. The average number of trimers per virion was estimated assuming there are 1400 Gag molecules present. External collaborators from the laboratory of Prof. Lifson (Frederick National Laboratory, USA) assisted this study by producing and purifying the viral Env

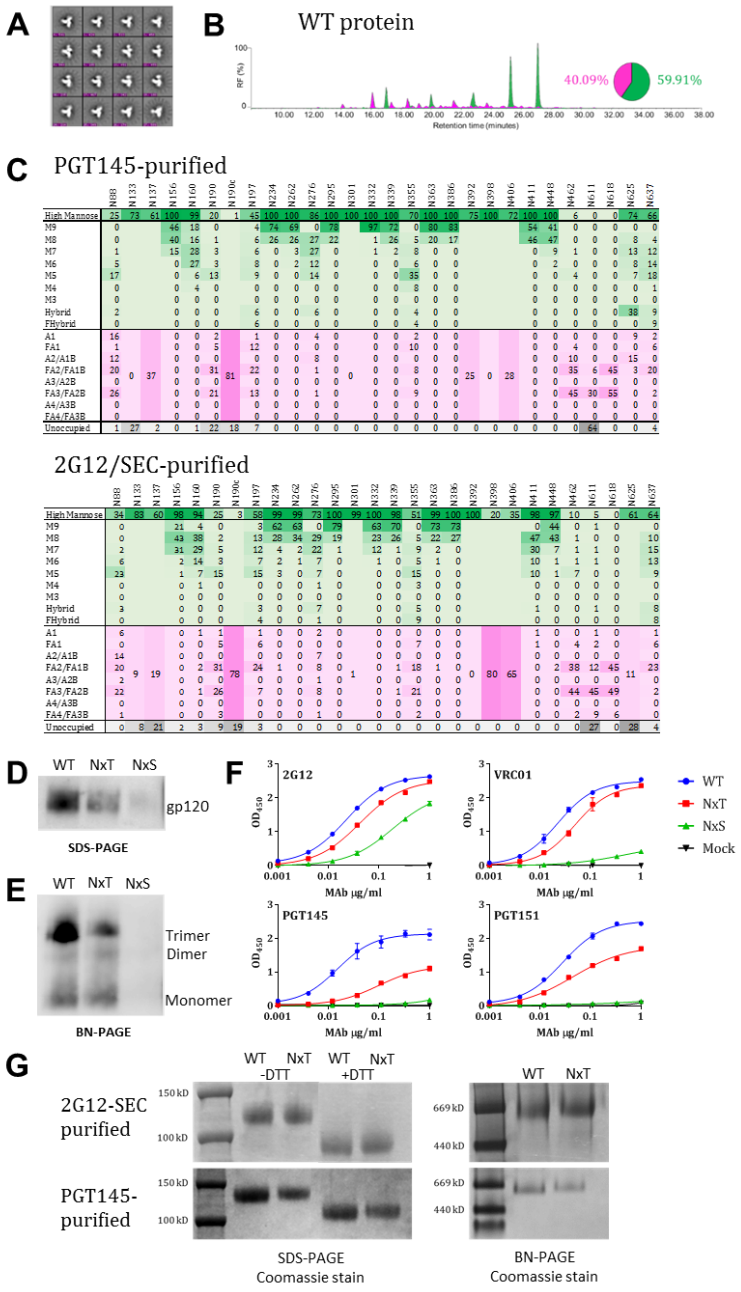
D.6 Quantification of N-linked glycans using exoglycosidase digests

		gp120			gp41		
		Viral-derived	SOSIP 293F	SOSIP CHO	Viral-derived	SOSIP 293F	SOSIP CHO
Oligomannose	M5	2.7	8.1	15	0.3	6.9	14
	M6	4.5	3.7	4	1.4	4.8	4
	M7	7.7	6.8	5	2.0	7.0	4
	M8	13.6	21.0	14	0.0	3.7	3
	M9	21.4	23.5	35	0.0	0.0	1
	TOTAL	49.9	63.1	73	3.7	22.4	26
Sialic Acids	α2-3	9.6	3.3	N.D.	9.3	4.2	N.D.
	α2-6	6.6	2.1	N.D.	16.9	3.4	N.D.
	TOTAL	16.3	5.3	N.D.	26.1	7.5	N.D.

Quantification of N-linked glycans using endoglycosidase H, α2,3 and α2,6 sialidase digests and subsequent HPLC analysis of gp120 and gp41 glycans from the BG505.T332N-LAI virion and 293F-derived BG505 SOSIP.664. Values represent percentages of total glycan population. CHO data reproduced from Dey *et al.* 2018. 'Sialic acids' refers to sialylated glycans.

Appendix E Supplementary information for: Enhancing glycan occupancy of soluble HIV-1 envelope trimers to mimic the native viral spike.

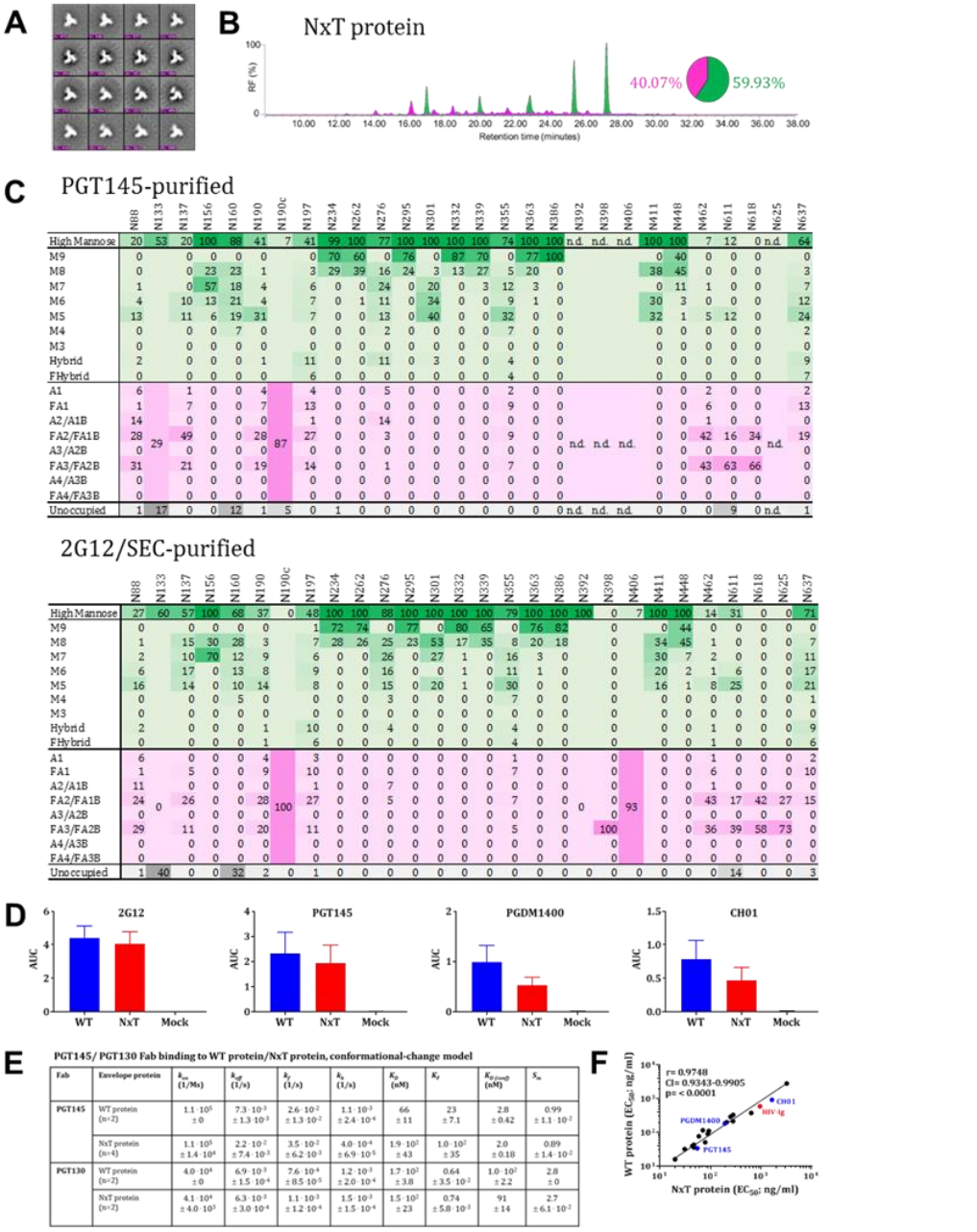
E.1 Several PNGS on BG505 SOSIP.v4.1 trimers are underoccupied.



(A) NS-EM analysis of 2G12/SEC-purified WT proteins, showing the 2D class-averages. (B) HILIC-UPLC analysis of the 2G12/SEC-purified WT protein (C) The data sets show the glycoforms found at each PNGS. Data for oligomannose/hybrid-type glycans are shaded in green, fully processed complex type glycans are shaded in magenta, while the absence of a glycan from some PNGS is shaded in grey. Oligomannose-type glycans are categorized according to the number of mannose residues present, hybrids by the presence/absence of fucose and complex-type glycans by the number of processed antenna and the

presence/absence of fucose. Data that could only be obtained from low intensity peptides cannot be allocated into the above categories. They are merged to cover all oligomannose/hybrid compositions or complex-type glycans. The data presented in this panel are the mean of two independent biological replicates of WT protein. (D) Reducing SDS-PAGE analysis of unpurified WT, NxT and NxS proteins expressed in 293T cells, followed by western blotting with the CA13 (ARP3119) mAb. (E) BN-PAGE analysis of the same proteins blotted with the 2G12 bNAb. The trimer, dimer and monomer bands are indicated. (F) D7324-capture ELISA quantifying the binding of the bNAbs 2G12, VRC01, PGT145 and PGT151 to the WT, NxT and NxS proteins. (G) Non-reducing and reducing (+ DTT) SDS-PAGE followed by Coomassie staining (left panel), and BN-PAGE analysis followed by Coomassie blue staining (right panel) of PGT145- and 2G12/SEC-purified WT and NxT proteins, as indicated. Electron microscopy was performed by Prof. Ward laboratory (Scripps Research Institute). Antigenic characterization was performed in the laboratories of Prof. Rogier Sanders (University of Amsterdam)

E.2 Glycan occupancy is enhanced by PNGS sequon engineering.



(A) NS-EM analysis of NxT proteins, showing the 2D class-averages. (B) HILIC-UPLC analysis of the NxT protein. The colour coding of the spectra and pie chart is the same as in Figure S1B. (C) Quantification of site-specific occupancy and composition for 28 PNGS on NxT trimers, purified using the 2G12/SEC and PGT145 methods as indicated. The data are derived from LC-ESI MS experiments. The data set shows the glycoforms found at each PNGS. The relative under-occupancy and oligomannose and complex/hybrid content at each individual site are summarized, using the same colour coding as in Figure S1C. (D) Binding of WT and NxT proteins to three V2-apex directed bNAbs, PGT145, PGDM1400 and CH01, and also 2G12 for comparison. AUC values derived from derived from ELISA titration curves are plotted. (E) Summary of the SPR binding kinetics of bNAbs PGT145 and PGT130. (F) The EC50 values derived using WT and NxT proteins were plotted and compared using Spearman's correlation coefficient. Binding data were generated for a panel of bNAbs spanning all major bNAb epitope clusters. All analyses were performed on NxT proteins produced in HEK293F followed by 2G12/SEC purification. Electron microscopy was performed by Prof. Ward laboratory (Scripps Research Institute). Antigenic

characterization was performed in the laboratories of Prof. Rogier Sanders (University of Amsterdam

E.3 Antigenic characterization of PGT145- and 2G12/SEC-purified WT and NxT proteins.

A

	Epitope	Antibody	Binding EC_{50} (ng/ml) PGT145 purified protein			Binding (AUC) PGT145 purified protein		
			WT protein	NxT protein	Fold difference	WT protein	NxT protein	Fold difference
Nabs	CD4bs	VRC01	66	100	1.5	3.2	2.9	0.9
		3BNC60	57	81	1.4	3.2	2.8	0.9
	V1/V2-glycan	PG9	388	359	0.9	3.0	2.7	0.9
		PG16	144	109	0.8	2.0	1.9	1.0
		PGT145	40	74	1.9	2.8	2.2	0.8
		PGDM1400	206	585	2.8	1.1	0.5	**0.5**
		CH01	372	906	2.4	2.2	1.4	**0.6**
	V3-glycan	PGT121	316	417	1.3	2.8	2.5	0.9
		PGT122	577	799	1.4	2.3	2.0	0.8
		PGT123	232	396	1.7	3.2	2.7	0.8
		PGT125	150	115	0.8	3.5	3.4	1.0
		PGT126	113	84	0.7	3.1	2.8	0.9
		PGT128	45	73	1.6	3.4	3.0	0.9
		PGT130	193	133	0.7	2.1	1.8	0.9
	OD-glycan	2G12	30	40	1.3	4.1	3.6	0.9
		PGT135	234	601	2.6	4.0	2.8	0.7
	gp120-gp41	PGT151	29	23	0.9	3.1	2.8	0.9
		35O22	814	2111	2.6	2.5	2.0	0.8
	gp41	3BC315	62	69	1.1	2.7	1.9	0.7
		HIV-IG	>10000	>10000	1.0	0.4	0.3	0.9

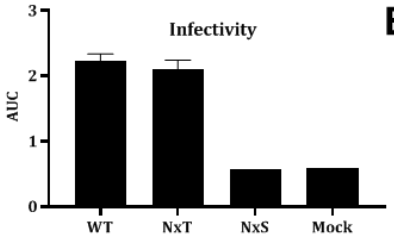
B

	Epitope	Antibody	Binding EC_{50} (ng/ml) 2G12/SEC purified protein			Binding (AUC) 2G12/SEC purified protein		
			WT protein	NxT protein	Fold difference	WT protein	NxT protein	Fold difference
Nabs	CD4bs	VRC01	36	47	1.3	3.6	3.3	0.9
		3BNC60	91	89	1.0	3.0	2.9	1.0
	V1/V2-glycan	PG9	201	211	1.1	2.9	3.1	1.1
		PG16	77	59	0.8	2.1	2.2	1.0
		PGT145	34	55	1.6	2.3	1.9	0.8
		PGDM1400	185	193	1.0	1.0	0.5	**0.5**
		CH01	923	1656	1.8	0.8	0.5	**0.6**
	V3-glycan	PGT121	291	250	0.9	2.3	2.4	1.0
		PGT122	333	282	0.8	2.0	1.7	0.9
		PGT123	216	280	1.3	2.2	1.8	0.9
		PGT125	51	79	1.5	3.5	3.3	0.9
		PGT126	32	31	1.0	4.1	4.1	1.0
		PGT128	43	47	1.1	4.1	3.8	0.9
		PGT130	111	93	0.8	2.0	1.8	0.9
	OD-glycan	2G12	16	20	1.3	4.4	4.0	0.9
		PGT135	379	644	1.7	3.7	2.7	0.8
	gp120-gp41	PGT151	40	44	1.1	1.8	1.5	0.8
		35O22	2829	3254	1.2	1.1	1.0	0.9
	gp41	3BC315	116	71	**0.6**	2.4	2.0	0.9
		HIV-IG	595	960	1.6	3.1	2.8	0.9

(A) Half-maximal binding concentrations (EC50; in ng/ml) were derived from D7324-capture ELISAs using PGT145-purified WT and NxT proteins. The values represent the means of 4–10 independent single titration experiments for each bNAb. The fold-differences in EC50 values for NxT vs. WT proteins are listed. Also tabulated are the average AUC values derived from the titration curves for each mAb. Bold numbers indicate values where the differences are <0.6- or >3-fold. (B) The same analysis and data presentation but derived using 2G12/SEC-purified WT and NxT proteins. Antigenic characterization was performed in the laboratories of Prof. Rogier Sanders (University of Amsterdam

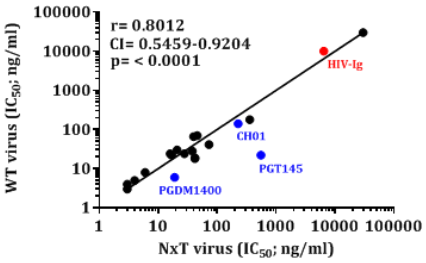
E.4 NxT sequon engineering is compatible with Env function and virus infectivity.

A



Virus	WT	NxT	NxS
CA-p24 (ng/ml)	127	1106	684

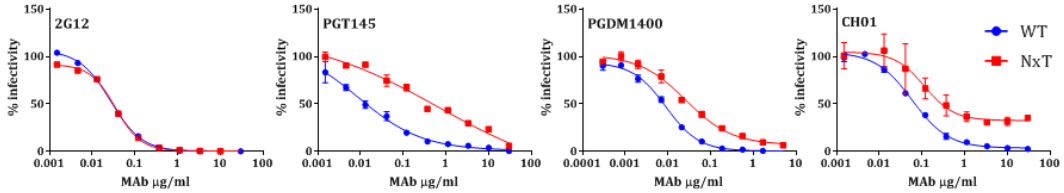
B



C

	Epitope	Antibody	Neutralization IC_{50} (ng/ml) WT virus	NxT virus	Fold difference	Neutralization (AUC) WT virus	NxT virus	Fold difference
Nabs	CD4bs	VRC01	66	40	0.6	238.2	200.7	0.8
		3BNC60	30	21	0.7	224.1	184.5	0.8
	V1/V2-glycan	PG9	19	42	2.2	155.0	181.2	1.2
		PG16	18	42	2.4	156.3	208.3	1.3
		PGT145	22	554	**25.7**	157.6	266.6	1.7
		PGDM1400	6	19	**3.1**	126.2	198.7	1.6
		CH01	140	227	1.6	203.8	279.1	1.4
	V3-glycan	PGT121	24	16	0.7	158.3	136.1	0.9
		PGT122	41	73	1.8	203.5	204.5	1.0
		PGT123	22	17	0.8	189.7	178.8	0.9
		PGT125	5	4	0.8	113.7	99.6	0.9
		PGT126	8	6	0.7	143.6	133.5	0.9
		PGT128	4	3	0.9	107.9	100.0	0.9
		PGT130	28	38	1.4	156.0	159.8	1.0
	OD-glycan	2G12	24	28	1.2	125.5	135.9	1.1
		PGT135	179	359	2.0	218.0	232.0	1.1
	gp120-gp41	PGT151	3	3	1.1	88.2	107.7	1.2
		35O22	70	46	0.6	144.7	117.8	0.8
	gp41	3BC315	>150000	>150000	1.0	359.3	372.1	1.0
		HIV-Ig	10086	6437	1.6	405.7	352.7	0.9

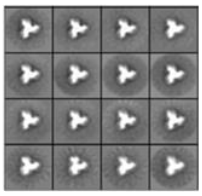
D



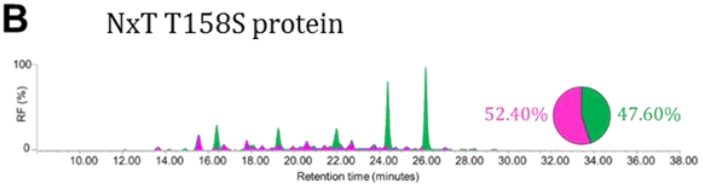
(A) Infection of TZM-bl reporter cells by WT, NxT and NxS viruses using a range of p24 concentrations (500, 250, 125, 62.5, 31.25, 16.62 and 7.81 pg). The AUC values derived from the titration curve were plotted. The box represents CA-p24 measured in each virus preparation. (B) Correlation between neutralization of WT and NxT viruses. The midpoint neutralization concentrations (IC50) were plotted and the Spearman's correlation coefficient, r, was calculated. The outliers, PGT145, PGDM1400 and CH01, are indicated in blue. Polyclonal HIV-Ig is indicated in red. (C) Midpoint neutralization concentrations (IC50; in ng/ml) were derived from single cycle infection experiments using TZM-bl cells and he indicated bNAbs or HIV-Ig. The values are averages based on 2-4 independent antibody-titration experiments. The average AUC values derived from the neutralization curves are also shown, in the three right-most columns. (D) Representative neutralization curves for the WT and NxT virus and the 2G12, PGT145, PGDM1400 and CH01 bNAbs.

E.5 N160 occupancy can be increased by reducing the affinity of a neighbouring site for OST.

A



B



C

PGT145-purified

	N88	N133	N137	N156	N160	N190	N190c	N197	N234	N262	N276	N295	N301	N332	N339	N355	N363	N386	N392	N398	N406	N411	N448	N462	N611	N618	N625	N637
High Mannose	11	65	17	100	100	25	4	38	100	100	63	100	98	100	100	64	100	100	100	n.d.	99	100	100	7	10	0	3	51
M9	0	1	0	2	0	0		0	69	69	0	77	0	81	67	0	78	69				0	33	0	0	0	0	0
M8	0	22	1	25	10	1		3	25	29	5	23	28	15	28	3	19	26				22	35	0	0	0	0	1
M7	1	15	1	39	19	4		5	3	2	6	0	39	4	5	9	2	5				33	20	1	2	0	0	4
M6	2	12	1	14	25	4		3	2	0	5	0	12	1	1	11	1	0				24	8	1	0	0	0	7
M5	6	13	1	16	31	13		10	0	0	5	0	14	0	0	24	0	0				19	3	4	8	0	3	11
M4	0	1	0	1	12	0		0	0	0	2	0	0	0	0	9	0	0				0	1	0	0	0	0	1
M3	0	0	0	0	3	0		0	0	0	0	0	0	0	0	0	0	0				0	0	0	0	0	0	0
Hybrid	2	2	9	1	0	2		8	0	0	38	0	3	0	0	6	0	0				2	0	1	0	0	0	11
FHybrid	1	0	5	0	0	1		7	0	0	2	0	1	0	0	2	0	0				0	0	1	0	0	0	15
A1	3	1	0	0	0	1		2	0	0	3	0	0	0	0	0	0	0				0	0	1	0	0	0	1
FA1	1	0	2	0	0	4		9	0	0	0	0	0	0	0	26	0	0				0	0	3	0	0	0	8
A2/A1B	7	0	4	0	0	0		1	0	0	26	0	0	0	0	0	0	0				0	0	5	0	0	1	0
FA2/FA1B	22	4	36	0	0	33	96	34	0	0	4	0	1	0	0	5	0	0	0	n.d.	1	0	0	38	10	25	29	31
A3/A2B	2	0	17	0	0	0		0	0	0	3	0	0	0	0	0	0	0				0	0	1	0	0	0	0
FA3/FA2B	39	4	20	0	0	32		15	0	0	0	0	1	0	0	4	0	0				0	0	40	59	45	46	6
A4/A3B	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0				0	0	0	0	0	0	0
FA4/FA3B	14	0	4	0	0	4		0	0	0	0	0	0	0	0	0	0	0				0	0	5	13	29	21	0
Unoccupied	1	26	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	n.d.	n.d.	0	0	0	8	0	0	2

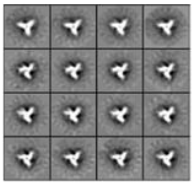
2G12/SEC-purified

	N88	N133	N137	N156	N160	N190	N190c	N197	N234	N262	N276	N295	N301	N332	N339	N355	N363	N386	N392	N398	N406	N411	N448	N462	N611	N618	N625	N637
High Mannose	17	75	n.d.	100	75	30	n.d.	40	95	100	31	100	n.d.	99	88	73	97	45	0	n.d.	n.d.	100	94	11	24	1	69	66
M9	0	0		20	2	0		1	61	65	0	63		78	56	0	73	32	0			0	37	0	0	0	0	0
M8	1	24		37	45	3		5	28	28	0	29		18	24	8	17	12	0			35	38	1	1	0	14	6
M7	2	19		30	15	6		5	4	3	1	3		2	5	14	3	0	0			32	10	2	4	0	13	9
M6	4	11		7	4	4		4	2	2	2	1		0	1	9	1	0	0			19	3	1	3	0	9	13
M5	9	17		5	3	13		10	1	1	4	2		0	0	26	1	0	0			13	3	6	13	0	7	16
M4	0	1		1	0	0		0	0	0	2	1		0	0	1	0	0	0			0	1	0	0	0	0	1
M3	0	0		0	0	0		0	0	0	0	0		0	0	0	0	0	0			0	0	0	0	0	0	0
Hybrid	2	2		0	5	2		7	0	0	20	0		0	0	6	0	0	0			2	1	1	2	0	25	10
FHybrid	1	1		0	1	2		7	0	0	3	0		0	1	8	0	1	0			0	1	1	1	0	2	11
A1	2	1		0	2	1		2	0	0	1	0		0	0	1	0	0	0			0	0	1	1	0	2	1
FA1	1	1		0	0	5		10	0	0	1	0		0	0	7	0	0	0			0	1	3	2	0	3	8
A2/A1B	8	0		0	0	0		1	0	0	13	0		0	0	0	0	2	0			0	0	1	0	0	10	0
FA2/FA1B	22	5	n.d.	0	15	30	n.d.	33	4	0	21	0	n.d.	0	5	11	2	23	2	n.d.	n.d.	0	4	43	13	22	7	19
A3/A2B	3	0		0	0	0		0	0	0	2	0		0	0	0	0	0	0			0	0	1	0	0	2	0
FA3/FA2B	34	4		0	7	29		13	1	0	25	0		0	6	7	1	27	65			0	1	36	44	46	6	5
A4/A3B	0	0		0	0	0		0	0	0	0	0		0	0	0	0	0	0			0	0	0	0	0	0	0
FA4/FA3B	11	0		0	0	4		1	0	0	4	0		0	1	1	0	4	33			0	0	5	12	31	0	1
Unoccupied	1	13	n.d.	0	0	1	n.d.	0	0	0	0	0	n.d.	0	0	0	0	0	0	n.d.	n.d.	0	0	0	3	0	0	0

(A) NS-EM analysis of NxT 158S trimers, showing the 2D class-averages. (B) HILIC-UPLC analysis of the NxT T158S protein. (C) Quantification of site-specific occupancy and composition for 28 PNGS on NxT T158S proteins, purified using the 2G12/SEC and PGT145 methods as indicated. The data are derived from LC-ESI MS experiments. The data set shows the glycoforms found at each PNGS. The relative under-occupancy and oligomannose and complex/hybrid content at each individual site are summarized. Electron microscopy was performed by Prof. Ward laboratory (Scripps Research Institute).

E.6 N133 occupancy is increased by reducing its affinity for OST.

A



B PGT145-purified

	N88	N133	N137	N156	N160	N190	N190c	N197	N234	N262	N276	N295	N301	N332	N339	N355	N363	N386	N392	N398	N406	N411	N448	N462	N611	N618	N625	N637
High Mannose	10	90	65	100	100	0	0	30	99	100	76	100	98	100	93	61	100	100	n.d.	n.d.	36	100	100	3	6	0	n.d.	52
M9	0			37	0			0	70	63	0	88		91	61	0	77	81				0	48	0	0	0		0
M8	0			46	33			1	24	29	10	12		8	27	3	21	19				45	39	0	0	0		0
M7	1			12	17			3	3	4	16	0		1	3	8	2	0				15	8	0	1	0		6
M6	2			3	20			3	2	4	15	0		0	2	7	1	0				26	4	0	1	0		5
M5	6			2	31			10	0	0	19	0		0	0	24	0	0				14	1	2	4	0		7
M4	0			0	0			0	0	0	2	0		0	0	1	0	0				0	0	0	0	0		1
M3	0			0	0			0	0	0	0	0		0	0	0	0	0				0	0	0	0	0		0
Hybrid	1			1	0			6	0	0	14	0		0	0	6	0	0				0	0	0	0	0		6
FHybrid	1			0	0			7	0	0	1	0		0	0	11	0	0				0	0	1	0	0		27
A1	3			0	0			2	0	0	5	0		0	0	1	0	0				0	0	0	0	0		1
FA1	1			0	0			11	0	0	0	0		0	0	15	0	0				0	0	4	1	0		10
A2/A1B	10			0	0			1	0	0	14	0		0	0	0	0	0				0	0	1	0	0		0
FA2/FA1B	29			0	0			38	0	0	2	0		0	0	12	0	0				0	0	43	14	28		27
A3/A2B	5	5	35	0	0	100	100	0	0	0	2	0	2	0	0	0	0	0	n.d.	n.d.	64	0	0	0	0	0	nd.	0
FA3/FA2B	35			0	0			16	0	0	0	0		0	0	9	0	0				0	0	43	55	60		9
A4/A3B	0			0	0			0	0	0	0	0		0	0	0	0	0				0	0	0	0	0		0
FA4/FA3B	8			0	0			1	0	0	0	0		0	0	1	0	0				0	0	5	20	10		0
Unoccupied	0	6	0	0	0	0	0	1	1	0	0	0	0	0	7	0	0	0	n.d.	n.d.	0	0	0	0	3	2	n.d.	2

2G12/SEC-purified

	N88	N133	N137	N156	N160	N190	N190c	N197	N234	N262	N276	N295	N301	N332	N339	N355	N363	N386	N392	N398	N406	N411	N448	N462	N611	N618	N625	N637
High Mannose	7	62	39	99	99	25	0	35	80	99	36	100	100	100	87	68	98	85	n.d.	n.d.	0	100	92	7	10	1	65	60
M9	0			24	9	0		1	52	66	0			82	64	0	77	68				0	39	0	0	0		0
M8	0			33	27	2		3	18	25	3			14	19	6	16	17				38	35	0	1	0		5
M7	0			27	21	3		6	4	4	6			2	3	13	3	0				29	11	1	2	0		7
M6	1			7	15	3		3	3	3	4			1	0	9	1	0				16	4	1	1	0		1
M5	5			6	19	17		11	2	2	4			2	0	26	1	0				17	2	4	6	0		29
M4	0			0	7	0		0	0	0	1			0	0	1	0	0				0	0	0	0	0		0
M3	0			0	1	0		0	0	0	0			0	0	0	0	0				0	0	0	0	0		0
Hybrid	0			1	0	1		6	0	0	15			0	0	6	0	0				0	1	0	0	0		7
FHybrid	1			0	0	0		5	1	0	2			0	1	6	0	0				0	1	1	1	1		10
A1	0			1	0	1		2	0	0	2			0	0	1	0	0				0	0	1	1	0		3
FA1	2			0	0	5		10	1	0	1			0	0	10	0	0				0	1	5	8	0		14
A2/A1B	8			0	0	0		1	0	0	11			0	0	0	0	0				0	0	2	0	0		0
FA2/FA1B	31			1	1	18		35	12	1	19			0	7	10	1	7				0	5	41	15	41		21
A3/A2B	3	0	60	0	0	0	100	0	0	0	4	0	0	0	0	0	0	0	nd.	nd.	91	0	0	0	0	0	21	0
FA3/FA2B	41			0	0	33		17	7	0	23			0	6	7	0	9				0	2	40	49	46		1
A4/A3B	0			0	0	0		0	0	0	0			0	0	0	0	0				0	0	0	0	0		0
FA4/FA3B	7			0	0	3		1	1	0	5			0	1	3	0	0				0	0	4	15	12		0
Unoccupied	0	37	1	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	n.d.	n.d.	0	0	0	0	0	0	13	1

C

	N88	N133	N137	N156	N160	N190	N190c	N197	N234	N262	N276	N295	N301	N332	N339	N355	N363	N386	N392	N398	N406	N411	N448	N462	N611	N618	N625	N637
CHO BG505 WT																												
High Mannose	2	83	69	92	100	0	0	19	100	100	99	100	100	100	92	68	100	100	n.d.	n.d.	0	99	100	0	0	n.d.	0	38
Complex	98	0	15	0	0	8	62	48	0	0	1	0	0	0	0	32	0	0	n.d.	n.d.	100	1	0	100	37	n.d.	0	45
Unoccupied	0	17	16	8	0	92	38	34	0	0	0	0	0	0	0	8	0	0	n.d.	n.d.	0	0	0	0	63	n.d.	100	16
CHO BG505 NxT																												
High Mannose	0	86	66	100	100	0	0	0	100	100	92	100	100	100	100	88	100	100	n.d.	n.d.	n.d.	n.d.	100	0	0	0	0	92
Complex	100	6	18	0	0	93	100	100	0	0	8	0	0	0	0	12	0	0	n.d.	n.d.	n.d.	n.d.	0	100	100	100	0	0
Unoccupied	0	8	16	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	n.d.	n.d.	n.d.	n.d.	0	0	0	0	100	8
Change in occupancy	0	8	1	8	0	85	38	34	0	0	0	0	0	0	8	0	0	0	n.d.	n.d.	n.d.	n.d.	0	0	63	n.d.	0	8

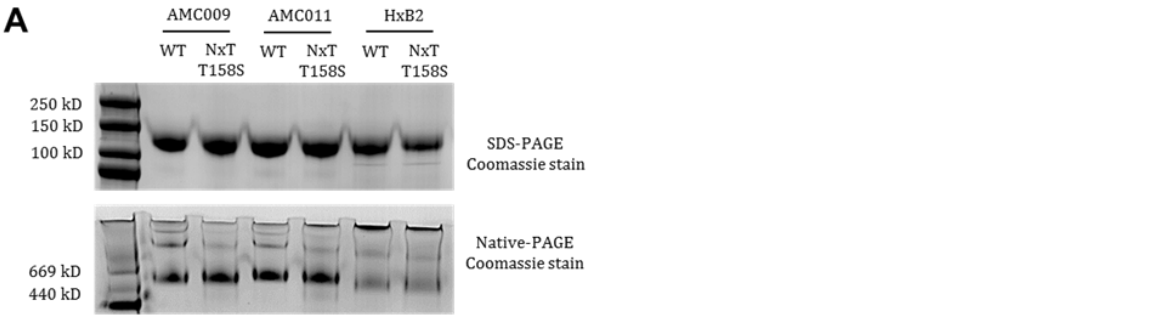
D

WT change in CHO	N88	N133	N137	N156	N160	N190	N190c	N197	N234	N262	N276	N295	N301	N332	N339	N355	N363	N386	N392	N398	N406	N411	N448	N462	N611	N618	N625	N637
High Mannose	-23	10	8	-8	1	-19	-1	-26	0	0	13	0	0	0	-8	-2	0	0	n.d.	n.d.	-72	-1	0	-6	0	n.d.	-74	-27
Complex	24	0	-23	0	0	-50	-19	0	0	0	-13	0	0	0	0	2	0	0	n.d.	n.d.	72	1	0	6	1	n.d.	-26	15
Unoccupied	-1	-10	14	8	-1	70	20	26	0	0	0	0	0	0	8	0	0	0	n.d.	n.d.	0	0	0	0	-1	n.d.	100	12

NxT change in CHO	N88	N133	N137	N156	N160	N190	N190c	N197	N234	N262	N276	N295	N301	N332	N339	N355	N363	N386	N392	N398	N406	N411	N448	N462	N611	N618	N625	N637
High Mannose	-10	-4	2	0	0	0	0	-30	1	0	16	0	2	0	7	27	0	0	n.d.	n.d.	n.d.	n.d.	0	-3	-6	0	n.d.	41
Complex	11	1	-17	0	0	-7	0	32	0	0	-16	0	-2	0	0	-27	0	0	n.d.	n.d.	n.d.	n.d.	0	3	9	2	n.d.	-46
Unoccupied	0	3	15	0	0	7	0	-1	-1	0	0	0	0	0	-7	0	0	0	n.d.	n.d.	n.d.	n.d.	0	0	-3	-2	n.d.	6

(A) NS-EM analysis of NxT T135S T158S proteins, showing the 2D class-averages. (B) Quantification of site-specific occupancy and composition for 28 PNGS on NxT T135S T158S proteins, purified using the 2G12/SEC and PGT145 methods as indicated. The data set shows the glycoforms found at each PNGS. The relative under-occupancy and oligomannose and complex/hybrid content at each individual site are summarized. (C) Site-specific analysis of the WT and NxT T135S T158S proteins produced in ExpiCHO-S cells. (D) Site-specific changes in glycosylation when BG505 SOSIP WT is produced in ExpiCHO-S cells compared to HEK293F cells. A percentage point (p.p.) increase represents a higher abundance of a particular glycoform on ExpiCHO-S cells. Electron microscopy was performed by Prof. Ward laboratory (Scripps Research Institute)

E.7 PNGS sequon engineering on diverse Env isolates.



D

AMC009 WT	N88	N130	N136	N139	N141	N156	N160	N189A	N197	N234	N241	N262	N276	N289	N295	N301	N332	N339	N356	N362	N386	N392	N397	N448	N463	N611	N616	N625	N637
High mannose	37	95	54	14	5	90	87	37	42	98	98	100	99	90	100	99	99	100	15	100	100	100	9	100	4	25	26	19	23
Complex	62	5	40	4	91	0	12	62	36	0	0	0	1	0	0	0	1	0	85	0	0	0	91	0	96	16	18	80	12
Unoccupied	1	0	6	82	3	10	1	1	23	2	2	1	0	10	0	1	0	0	0	0	0	0	0	0	0	58	55	1	65
NxT																													
High mannose	37	100	70	6	3	88	92	20	22	98	97	100	93	100	100	100	99	100	0	100	100	100	0	100	0	1	2	18	12
Complex	55	0	26	85	94	0	3	75	73	0	2	0	7	0	0	0	0	0	100	0	0	0	100	0	100	84	94	82	18
Unoccupied	8	0	4	9	2	12	5	4	5	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	14	4	0	70
Change in occupancy	-7	0	2	73	1	-2	-4	-4	18	1	1	1	0	10	0	1	-1	0	0	0	0	0	0	0	0	44	51	1	-5

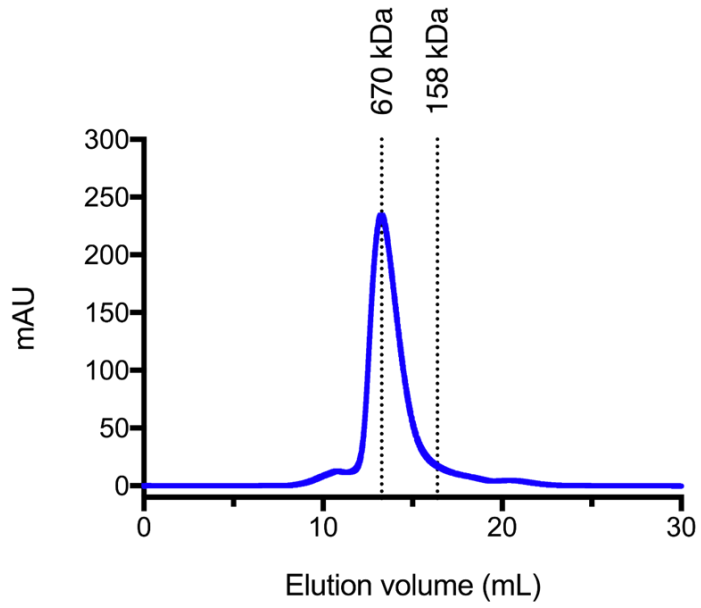
AMC011 WT	N88	N130	N136	N141	N141c	N156	N160	N188	N197	N234	N241	N262	N276	N289	N295	N301	N332	N339	N355	N362	N386	N392	N396	N406	N448	N461	N611	N616	N625	N637
High mannose	34	100	100	14	21	99	99	73	57	98	86	99	97	n.d.	99	99	100	100	22	97	100	n.d.	n.d.	100	98	36	3	3	19	44
Complex	65	0	0	8	79	0	1	26	41	0	0	0	3	n.d.	0	0	0	0	77	0	0	n.d.	n.d.	0	0	63	5	0	75	41
Unoccupied	1	0	0	78	0	1	0	1	2	2	14	1	0	n.d.	1	1	0	0	0	3	0	n.d.	n.d.	0	2	1	92	97	6	15
NxT																														
High mannose	50	96	100	18	0	88	99	89	81	99	99	100	29	99	91	95	100	87	23	100	100	100	100	99	100	25	9	14	57	35
Complex	41	1	0	82	79	0	0	8	19	0	1	0	71	0	0	5	0	0	77	0	0	0	0	1	0	75	88	6	39	27
Unoccupied	8	3	0	0	21	12	0	2	0	0	0	0	0	1	9	0	0	13	0	0	0	0	0	0	0	0	3	80	4	38
Change in occupancy	-7	-3	0	78	-21	-12	0	-2	2	2	14	1	0	n.d.	-8	1	0	-13	0	3	0	n.d.	n.d.	0	2	1	89	17	2	-23

HXB2 WT	N88	N136	N141	N156	N160	N189	N197	N230	N234	N241	N262	N276	N289	N295	N301	N322	N339	N356	N386	N392	N397	N406	N411	N448	N463	N611	N616	N624	N637
High Mannose	94	12	0	52	20	6	34	100	96	97	100	75	98	100	100	98	99	25	100	99	n.d.	7	n.d.	98	25	3	2	56	20
Complex	6	88	100	3	3	33	60	0	1	1	0	14	0	0	0	0	0	75	0	1	n.d.	93	n.d.	1	73	43	15	40	39
Unoccupied	0	0	0	45	77	61	5	0	2	2	0	11	1	0	0	1	0	0	0	0	n.d.	0	n.d.	0	2	54	84	3	41
NxT																													
High Mannose	86	14	3	54	66	5	30	100	97	96	100	77	100	99	28	95	97	11	99	99	100	0	100	94	11	2	14	58	20
Complex	14	86	97	13	22	42	69	0	3	3	0	20	0	1	70	5	3	89	1	1	0	100	0	6	89	89	68	42	64
Unoccupied	0	0	0	33	11	53	1	0	0	0	0	3	0	0	2	0	0	0	0	0	0	0	0	0	0	9	18	0	16
Change in occupancy	0	0	0	12	66	8	5	0	2	2	0	8	1	0	-2	1	0	0	0	0	n.d.	0	n.d.	0	2	46	65	3	25

(A) Reducing (+ DTT) SDS-PAGE followed by Coomassie staining (top panel), and BN-PAGE analysis followed by Coomassie blue staining (bottom panel) of PGT145-purified AMC009, AMC011 and HxB2 WT and NxT T158S proteins, as indicated. (B) D7324-capture ELISA quantifying the binding of the bNAbs 2G12, VRC01, PGT145 and PGT151 to the AMC009, AMC009 NxT T158S, AMC011 and AMC011 NxT T158S proteins. (C) Ni-NTA-capture ELISA quantifying the binding of the bNAbs 2G12, VRC01, PGT145 and PGT151 to the HxB2 and HxB2 NxT T158S proteins. (D) Glycoforms are grouped for all samples into high mannose (oligomannose- and hybrid-type), complex and unoccupied. The percentage change in occupancy between the WT protein and the NxT T158S protein is also shown. Electron microscopy was performed by Prof. Ward laboratory (Scripps Research Institute). Antigenic characterization was performed in the laboratories of Prof. Rogier Sanders (University of Amsterdam).

Appendix F Supplemental information for Site-specific glycan analysis of the SARS-CoV-2 spike

F.1 Size-exclusion chromatogram of the affinity purified SARS-CoV-2 S protein.



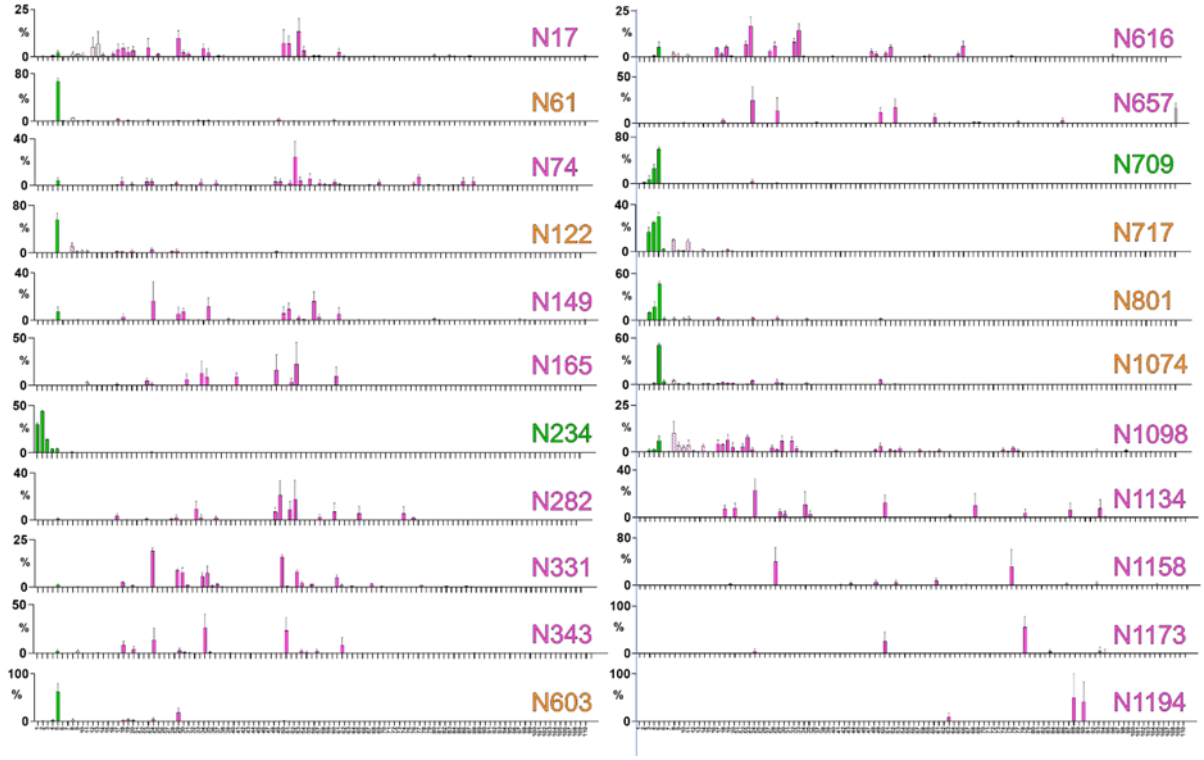
The elution volume of 670 kDa corresponding to the trimeric mass of the protein is shown.

F.2 Glycoform abundances observed across SARS CoV-2 S protein.

	Glycan	17	61	74	122	149	165	234	282	331	343	603	616	657	709	717	801	1074	1098	1134	1158	1173	1194
1	HexNAc(2)Hex(9)	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	HexNAc(2)Hex(8)	0	0	0	0	0	0	44	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0
3	HexNAc(2)Hex(7)	0	0	0	0	0	0	14	0	0	0	0	0	0	7	17	10	0	1	0	0	0	0
4	HexNAc(2)Hex(6)	1	1	0	0	0	0	4	0	0	0	2	1	0	26	25	17	2	2	0	0	0	0
5	HexNAc(2)Hex(5)	2	67	4	56	7	0	4	1	1	2	62	5	0	59	30	47	51	6	0	0	1	0
6	HexNAc(2)Hex(4)	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2	2	5	0	0	0	0	0
7	HexNAc(2)Hex(4)Fuc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	HexNAc(3)Hex(5)	1	6	0	11	0	0	1	0	0	0	3	2	0	0	9	2	5	10	0	0	0	0
9	HexNAc(3)Hex(5)Fuc(1)	1	1	0	2	0	0	0	0	0	2	0	1	0	0	1	0	1	4	0	0	0	0
10	HexNAc(3)Hex(5)NeuAc(1)	1	1	0	3	0	0	0	0	0	0	0	0	0	0	1	2	0	2	0	0	0	0
11	HexNAc(3)Hex(6)	0	1	0	3	0	3	0	0	0	0	0	1	0	0	9	3	1	4	0	0	0	0
12	HexNAc(3)Hex(6)Fuc(1)	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
13	HexNAc(3)Hex(6)Fuc(1)NeuAc(1)	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	HexNAc(3)Hex(6)NeuAc(1)	1	0	0	1	0	0	0	0	0	0	0	0	0	0	2	0	1	3	0	0	0	0
15	HexNAc(2)Hex(3)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
16	HexNAc(2)Hex(3)Fuc(1)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	HexNAc(3)Hex(3)	3	4	1	3	0	1	0	3	0	0	0	5	0	0	0	3	2	4	0	0	0	0
18	HexNAc(3)Hex(3)Fuc(1)	5	0	3	2	3	0	0	0	3	8	3	1	3	0	0	1	3	4	7	0	0	0
19	HexNAc(3)Hex(4)	2	2	0	0	0	0	0	0	0	0	4	5	0	0	2	0	2	6	0	1	0	0
20	HexNAc(3)Hex(4)Fuc(1)	3	1	1	3	0	0	0	0	1	4	2	1	0	0	1	0	2	2	8	0	0	0
21	HexNAc(3)Hex(4)Fuc(1)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
22	HexNAc(3)Hex(4)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
23	HexNAc(4)Hex(3)	5	2	3	0	0	5	0	1	0	0	0	7	0	0	0	0	1	8	0	0	0	0
24	HexNAc(4)Hex(3)Fuc(1)	0	0	3	5	16	2	1	0	19	14	4	17	25	4	0	3	5	1	23	0	4	0
25	HexNAc(4)Hex(3)Fuc(2)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	HexNAc(4)Hex(3)Fuc(3)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	HexNAc(4)Hex(3)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	HexNAc(4)Hex(4)	0	1	1	2	0	0	0	1	0	0	0	3	0	0	0	0	0	2	0	40	0	0
29	HexNAc(4)Hex(4)Fuc(1)	10	1	2	3	6	0	0	2	9	3	19	6	14	1	0	3	3	1	5	0	0	0
30	HexNAc(4)Hex(4)Fuc(1)NeuAc(1)	2	0	0	0	7	0	0	0	8	1	0	0	0	0	0	0	1	6	3	0	0	0
31	HexNAc(4)Hex(4)Fuc(2)	1	0	0	0	0	6	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
32	HexNAc(4)Hex(4)NeuAc(1)	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0
33	HexNAc(4)Hex(5)	0	1	0	0	0	0	0	9	0	0	0	8	0	0	0	0	0	2	0	0	0	0
34	HexNAc(4)Hex(5)Fuc(1)	4	1	3	1	0	13	0	2	6	27	0	14	0	0	0	0	0	0	11	0	0	0
35	HexNAc(4)Hex(5)Fuc(1)NeuAc(1)	2	1	0	1	11	9	0	0	7	1	0	0	0	0	0	2	2	0	3	0	0	0
36	HexNAc(4)Hex(5)Fuc(1)NeuAc(2)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
37	HexNAc(4)Hex(5)Fuc(2)	1	1	2	0	0	0	0	2	2	0	0	0	1	0	0	0	0	0	0	0	0	0
38	HexNAc(4)Hex(5)Fuc(2)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	HexNAc(4)Hex(5)Fuc(3)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40	HexNAc(4)Hex(5)Fuc(3)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41	HexNAc(4)Hex(5)NeuAc(1)	0	1	0	1	0	9	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
42	HexNAc(4)Hex(5)NeuAc(2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	HexNAc(4)Hex(6)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
44	HexNAc(4)Hex(6)Fuc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45	HexNAc(4)Hex(6)Fuc(1)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
46	HexNAc(4)Hex(6)Fuc(2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47	HexNAc(4)Hex(6)Fuc(3)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48	HexNAc(4)Hex(6)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0
49	HexNAc(5)Hex(3)	0	3	3	2	0	16	0	7	0	0	0	3	0	0	0	0	0	1	0	0	0	0
50	HexNAc(5)Hex(3)Fuc(1)	7	0	3	1	6	0	0	21	16	23	1	1	11	1	0	2	6	3	12	0	25	0
51	HexNAc(5)Hex(3)Fuc(1)NeuAc(1)	7	0	0	0	9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
52	HexNAc(5)Hex(4)	0	1	2	1	0	4	0	9	0	0	0	2	0	0	0	0	0	1	0	4	0	0
53	HexNAc(5)Hex(4)Fuc(1)	13	0	24	0	2	23	0	18	8	2	0	5	17	0	0	0	1	1	0	0	0	0
54	HexNAc(5)Hex(4)Fuc(1)NeuAc(1)	3	0	4	0	1	0	0	0	2	1	0	0	0	0	0	0	0	2	0	0	0	0
55	HexNAc(5)Hex(4)Fuc(1)NeuAc(2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
56	HexNAc(5)Hex(4)Fuc(2)	0	0	6	0	16	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0
57	HexNAc(5)Hex(4)Fuc(2)NeuAc(1)	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58	HexNAc(5)Hex(4)NeuAc(1)	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0
59	HexNAc(5)Hex(4)NeuAc(2)	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60	HexNAc(5)Hex(5)	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0
61	HexNAc(5)Hex(5)Fuc(1)	2	0	3	0	5	10	0	7	5	8	0	1	6	0	0	0	0	0	0	0	0	0
62	HexNAc(5)Hex(5)Fuc(1)NeuAc(1)	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
63	HexNAc(5)Hex(5)Fuc(1)NeuAc(2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	8
64	HexNAc(5)Hex(5)Fuc(2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
65	HexNAc(5)Hex(5)Fuc(3)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
66	HexNAc(5)Hex(5)NeuAc(1)	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	1	0	0
67	HexNAc(5)Hex(6)	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
68	HexNAc(5)Hex(6)Fuc(1)	0	0	1	0	0	0	0	0	2	0	0	6	0	0	0	0	0	0	10	0	0	0
69	HexNAc(5)Hex(6)Fuc(1)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
70	HexNAc(5)Hex(6)Fuc(1)NeuAc(2)	0	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
71	HexNAc(5)Hex(6)Fuc(1)NeuAc(3)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
72	HexNAc(5)Hex(6)Fuc(2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
73	HexNAc(5)Hex(6)Fuc(2)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
74	HexNAc(5)Hex(6)Fuc(3)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
75	HexNAc(5)Hex(6)NeuAc(1)	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	1	0	32	0	0
76	HexNAc(5)Hex(6)NeuAc(2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
77	HexNAc(6)Hex(3)	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0	0
78	HexNAc(6)Hex(3)Fuc(1)	0	0	7	0	0	0	0	0	1	0	0	1	1	0	0	0	0	1	4	0	55	0
79	HexNAc(6)Hex(3)Fuc(1)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
80	HexNAc(6)Hex(3)Fuc(2)	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
81	HexNAc(6)Hex(3)Fuc(3)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
82	HexNAc(6)Hex(4)	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
83	HexNAc(6)Hex(4)Fuc(1)	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	4	0	0
84	HexNAc(6)Hex(4)Fuc(2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
85	HexNAc(6)Hex(4)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
86	HexNAc(6)Hex(5)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
87	HexNAc(6)Hex(5)Fuc(1)	0	0	4	0	0	0	0	0	1	0	0	0	3	0	0	0	0	0	6	0	0	0
88	HexNAc(6)Hex(5)Fuc(1)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50
89	HexNAc(6)Hex(5)Fuc(1)NeuAc(2)	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
90	HexNAc(6)Hex(5)Fuc(1)NeuAc(3)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	42
91	HexNAc(6)Hex(5)Fuc(2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
92	HexNAc(6)Hex(6)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
93	HexNAc(6)Hex(6)Fuc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	6	0
94	HexNAc(6)Hex(6)Fuc(1)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	4	0
95	HexNAc(6)Hex(6)Fuc(1)NeuAc(2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
96	HexNAc(6)Hex(6)Fuc(2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
97	HexNAc(6)Hex(6)Fuc(3)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
98	HexNAc(6)Hex(7)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
99	HexNAc(6)Hex(7)Fuc(1)	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
100	HexNAc(6)Hex(7)Fuc(1)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
101	HexNAc(6)Hex(7)Fuc(1)NeuAc(2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
102	HexNAc(6)Hex(7)Fuc(1)NeuAc(3)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
103	HexNAc(6)Hex(7)Fuc(4)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
104	HexNAc(6)Hex(7)NeuAc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
105	HexNAc(6)Hex(7)NeuAc(2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
106	HexNAc(6)Hex(7)NeuAc(4)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
107	HexNAc(7)Hex(3)Fuc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
108	HexNAc(7)Hex(4)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
109	HexNAc(7)Hex(4)Fuc(1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
110	Unoccupied	0	0	0	0	1	0	0	0	0	0	0	0	16	0	0	0	0	0	0	0	0	0

The average abundance of glycan compositions across three biological replicates is shown, with a white to red colour scale used to highlight higher abundances of particular glycan compositions.

F.3 Extended site-specific N-linked glycosylation of SARS-CoV-2 S glycoprotein.



All glycan compositions detected across every N-linked glycan site are listed and numbered along the x-axis. The corresponding glycan compositions can be found in supplementary table 2. The bar graphs represent the mean abundance of each glycan of three biological repeats (+/- SEM) with oligomannose-type glycan series (green), hybrid glycans (dashed pink), and complex glycans (pink). Glycan sites are coloured according to oligomannose-type glycan content with the glycan sites labelled in green (80−100%), orange (30−79%) and pink (0−29%).

F.4 Glycoform abundances observed across SARS CoV-2 S protein

	N17	N61	N74	N122	N149	N165	N234	N282	N331	N343	N603	N616	N657	N709	N717	N801	N1074	N1098	N1134	N1158	N1173	N1194
M9	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M8	0	0	0	0	0	0	44	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0
M7	0	0	0	0	0	0	14	0	0	0	0	0	0	7	17	10	0	1	0	0	0	0
M6	0	1	0	0	0	0	4	0	0	0	2	1	0	26	25	17	2	2	0	0	0	0
M5	3	68	4	56	7	0	4	1	1	2	62	5	0	59	32	49	55	6	0	0	1	0
Hybrid	0	8	0	17	0	3	1	0	0	0	3	3	0	0	21	7	8	19	0	0	0	0
Fhybrid	2	1	0	2	0	1	0	0	0	2	0	1	0	0	1	0	1	5	0	0	0	0
A1	0	5	1	3	0	0	0	4	0	0	4	10	0	0	2	3	5	14	0	1	0	0
FA1	6	1	5	5	3	0	1	0	4	12	4	2	3	0	1	1	5	7	15	0	0	0
A2/A1B	0	6	5	3	0	14	0	11	0	0	0	18	0	0	0	1	2	19	0	48	0	0
FA2/FA1B	32	4	11	10	42	29	1	6	53	46	23	37	39	5	1	8	12	9	44	0	4	0
A3/A2B	0	6	10	3	0	20	0	30	0	0	0	6	0	0	0	0	0	5	0	44	0	0
FA3/FA2B	53	1	45	1	42	32	0	46	38	37	1	13	37	1	0	2	8	7	23	0	25	8
A4/A3B	0	0	3	0	0	0	0	2	0	0	0	0	0	0	0	0	0	3	0	6	0	0
FA4/FA3B	3	0	16	0	3	0	0	0	3	0	0	2	4	0	0	0	1	4	18	0	70	92
Unoccupied	1	0	0	0	1	0	0	0	0	0	0	0	16	0	0	0	0	0	0	0	0	0

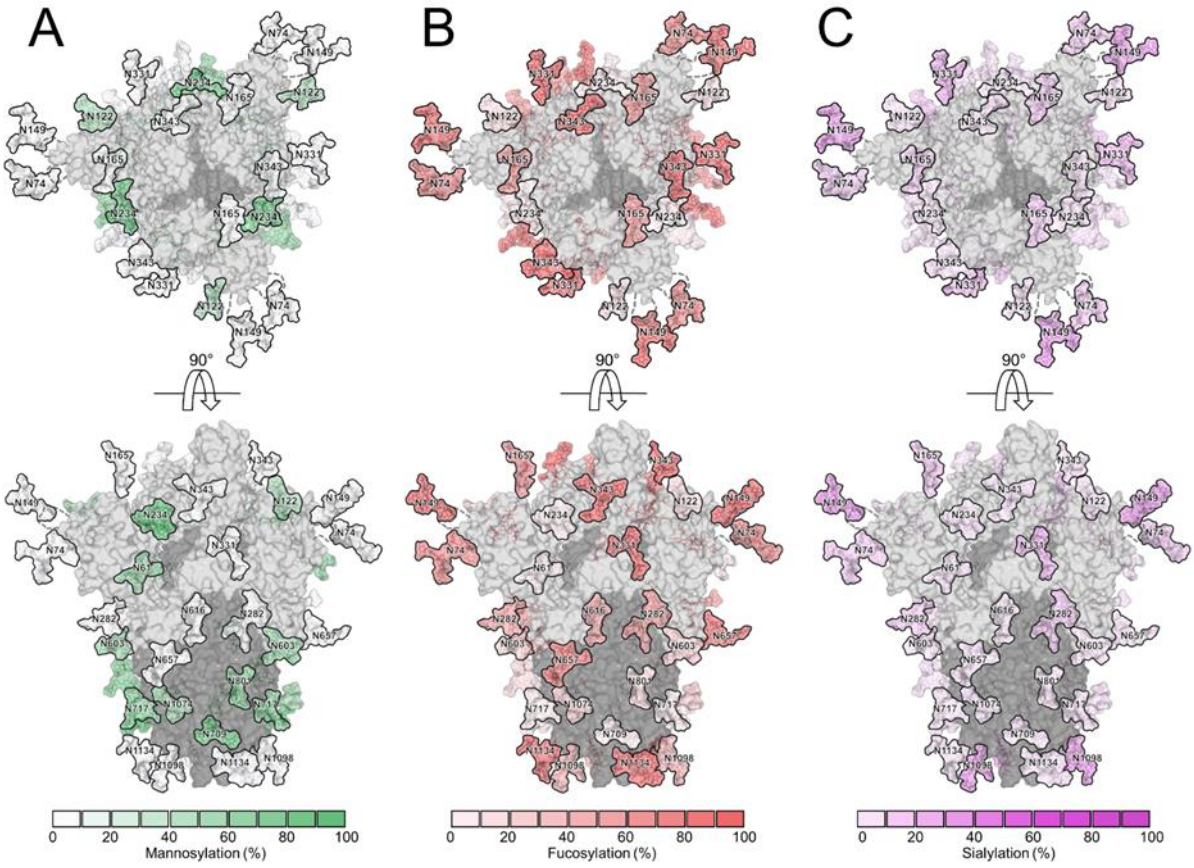
Total	
M9	1
M8	2
M7	2
M6	4
M5	19
Hybrid	4
Fhybrid	1
A1	2
FA1	3
A2/A1B	6
FA2/FA1B	19
A3/A2B	6
FA3/FA2B	19
A4/A3B	1
FA4/FA3B	10
Unoccupied	1

%	N17	N61	N74	N122	N149	N165	N234	N282	N331	N343	N603	N616	N657	N709	N717	N801	N1074	N1098	N1134	N1158	N1173	N1194
Mannose	4	69	4	56	8	0	97	1	2	2	65	6	0	94	74	77	57	9	0	0	1	0
Hybrid	2	9	0	19	1	5	1	0	0	3	3	4	0	0	22	7	9	24	0	0	0	0
Complex	94	22	96	25	91	95	2	99	98	96	32	90	83	6	4	16	33	67	100	100	99	100
Unoccupied	1	0	0	0	1	0	0	0	0	0	0	0	16	0	0	0	0	0	0	0	0	0
Fucosylation	95	6	77	18	91	63	2	52	98	98	29	56	83	5	3	11	27	31	100	0	99	100
Sialylation	20	4	18	5	33	18	0	14	22	4	0	1	2	0	2	5	6	30	7	39	4	100

Total	
Mannose	28
Hybrid	5
Complex	66
Unoccupied	1
Fucosylation	52
Sialylation	15

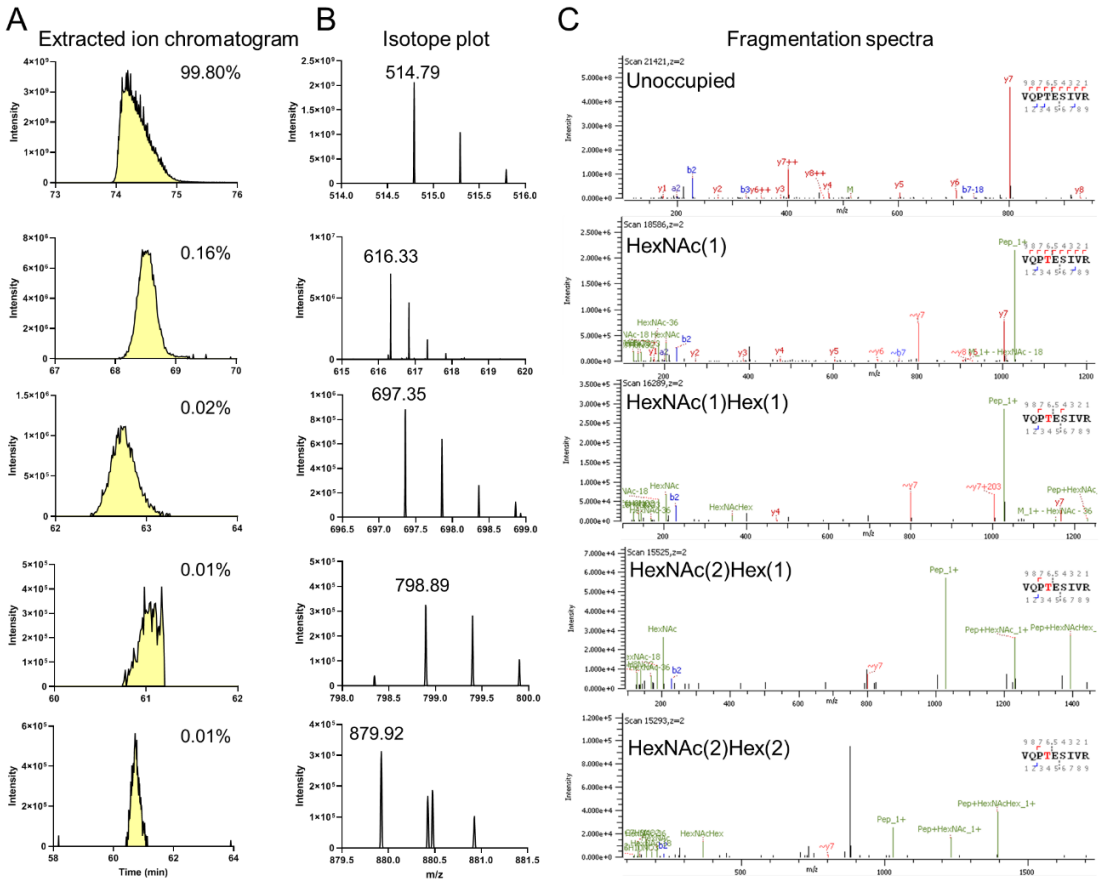
The upper table shows the categorized glycan compositions at each N-linked glycan site with the reported value the mean of three biological replicates. The global averages are shown in the right-hand table. The lower table further categorizes the glycan compositions into oligomannose-, hybrid-, and complex-type as well as the percentage of glycan compositions containing at least one fucose or one sialic acid residue.

F.5 Glycosylated model of SARS-CoV-2 S glycoprotein highlighting different glycan modifications.



The experimentally determined quantities of mannosylation, fucosylation and sialylation were used to colour the glycosylated model of SARS-CoV2 S protein. Glycans are highlighted according to mannosylation (A), fucosylation (B) and sialylation (C) levels as denoted in the keys. S1 and S2 subunits are coloured light grey and dark grey, respectively.

F.6 Detection of low levels of mucin-type O-linked glycosylation at T323/S325 of SARS-CoV-2 S



O-linked glycan compositions were observed at T323/S325. This analysis was performed on a single biological replicate. For each peptide/glycopeptide detected the extracted ion chromatogram (XIC) (A), isotope distribution (B), and fragmentation spectrum (C) is shown. The monoisotopic peak *m/z* is labelled in (B). Fragment ions are coloured blue and red for b- and y-ions, respectively. Oxonium ions are coloured in green.

Appendix G Supplementary information for Site-specific steric control of SARS-CoV-2 Spike glycosylation

G.1 Average recombinant and viral-derived S protein site-specific glycan compositions averaged

	N17	N61	N74	N122	N149	N165	N234	N282	N331	N343	N603	N616	N657	N709	N717	N801	N1074	N1098	N1134	N1158	N1173	N1194	Average
M9Glc	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0
M9	0	0	0	0	0	0	5	0	0	0	0	0		0	-1	-1	0	-1	0	0	0	0	0
M8	0	-2	0	0	0	0	25	0	0	0	3	0		-3	-1	-2	4	0	0	0	0	0	1
M7	0	10	0	-3	0	0	0	0	0	0	6	0		-8	8	7	8	0	0	0	0	0	1
M6	0	31	0	3	-1	-1	-2	0	0	0	10	-2		32	7	29	18	0	0	0	0	0	6
M5	-4	-14	-1	-12	-3	-17	-7	-3	-4	-2	-11	-2		-10	-3	-16	-14	-3	-3	-13	-7	-6	-7
M4	0	1	0	0	-2	0	-1	0	0	0	-2	0		0	-3	-3	0	0	0	0	0	0	-1
M3	0	-1	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0
FM	0	0	-1	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0
Hybrid	-1	0	0	24	0	-2	-4	1	1	0	14	12		6	22	16	14	14	-1	0	3	0	6
Fhybrid	-2	-1	-1	5	-1	-1	-1	0	-1	12	-3	-3		-4	-6	-1	-5	-1	12	0	-2	0	0
HexNAc(3)(x)	0	-5	0	0	0	-1	-1	1	1	0	-2	-4		0	-7	-4	-3	0	0	0	0	-1	-1
HexNAc(3)(F)(x)	-4	0	-3	-4	-4	0	0	1	0	-2	3	-4		-4	-3	-1	-3	-1	-2	0	-4	-7	-2
HexNAc(4)(x)	0	3	0	0	-2	95	-7	21	13	-1	18	39		14	-2	9	14	-8	0	0	12	0	10
HexNAc(4)(F)(x)	-6	-9	5	-1	14	-32	-1	12	-5	-54	-15	-3		-19	-8	-24	-4	-13	47	-16	7	-16	-7
HexNAc(5)(x)	8	-11	0	-3	-1	-4	0	-5	7	11	0	-6		0	0	-2	5	24	0	0	18	-13	1
HexNAc(5)(F)(x)	20	3	3	-6	8	-33	0	-22	-11	29	-22	-21		0	-2	-6	-24	-6	-20	-23	-2	97	-2
HexNAc(6+)(x)	0	-1	0	0	0	0	-6	-1	0	9	0	0		0	0	0	0	-1	0	0	0	0	0
HexNAc(6+)(F)(x)	-5	-2	-3	0	-6	-3	0	-3	-2	-1	0	-5		-3	0	0	-8	-5	-15	-47	11	-18	-5
Unoccupied	-6	0	0	-1	-1	0	-1	0	0	0	0	0		0	0	0	-2	0	-15	0	-35	-36	-5
Core	0	-1	0	0	0	0	0	0	1	0	0	0		0	0	-1	0	0	0	0	-1	0	0
Oligomannose	-5	24	-2	-12	-6	-19	20	-4	-4	-2	6	-4		11	7	13	16	-3	-5	-13	-7	-7	0
Hybrid	-3	-1	-1	28	-1	-2	-4	1	0	12	11	9		2	16	15	9	14	11	0	1	0	6
Complex	14	-23	3	-15	8	21	-15	3	2	-10	-17	-4		-13	-23	-27	-22	-10	9	-87	41	43	-6
Unoccupied	-6	0	0	-1	-1	0	-1	0	0	0	0	0		0	0	0	-2	0	-15	0	-35	-36	-5
Fucose	5	-9	2	-7	10	-69	-3	-12	-19	-16	-36	-37		-30	-19	-32	-43	-26	22	-86	11	57	-16
NeuAc	-26	-7	-43	-21	-35	-18	0	-15	-30	-37	-4	-9		-6	-6	-28	-13	-45	-18	-41	-14	-32	-21
Sulfation	0	0	-8	0	-1	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0

G.2 Percentage point change in glycosylation on viral-derived S protein compared to recombinant S protein.

Averaged	N17	N61	N74	N122	N149	N165	N234	N282	N331	N343	N603	N616	N657	N709	N717	N801	N1074	N1098	N1134	N1158	N1173	N1194
M9Glc	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M9	0	0	0	0	0	0	22	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0
M8	0	2	0	0	0	0	28	0	0	0	1	0	0	3	1	2	1	1	0	0	0	0
M7	0	2	0	3	0	0	15	0	0	0	2	0	0	8	14	5	1	2	0	0	0	0
M6	0	6	0	1	1	1	5	0	0	0	3	2	0	12	13	3	2	1	0	0	0	0
M5	4	26	1	12	3	17	7	3	4	2	17	2	3	26	17	16	14	4	3	13	7	6
M4	0	2	0	0	2	0	1	0	0	0	2	0	0	0	3	3	0	0	0	0	0	0
M3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FM	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hybrid	1	11	0	13	0	2	4	1	1	0	7	4	0	0	22	17	5	8	1	0	0	0
Fhybrid	2	1	1	3	1	1	1	1	1	1	4	3	1	4	6	1	5	2	2	0	2	0
HexNAc(3)(x)	0	5	0	7	0	1	1	1	0	0	5	4	0	0	7	4	3	3	0	0	0	1
HexNAc(3)(F)(x)	4	0	4	4	4	0	0	1	2	2	2	4	1	4	3	1	6	2	2	0	4	7
HexNAc(4)(x)	0	18	0	25	2	5	7	4	4	1	1	8	0	0	3	14	3	24	0	0	0	0
HexNAc(4)(F)(x)	40	9	30	18	43	32	1	30	52	58	34	39	43	24	8	24	24	15	39	16	15	16
HexNAc(5)(x)	0	11	0	5	1	4	0	10	0	0	0	6	1	0	0	2	1	13	0	0	0	13
HexNAc(5)(F)(x)	37	2	44	6	36	33	0	38	31	34	22	21	33	16	2	6	24	12	20	23	9	3
HexNAc(6+)(x)	0	1	0	0	0	0	6	1	0	0	0	0	0	0	0	0	0	8	0	0	0	0
HexNAc(6+)(F)(x)	5	2	19	0	6	3	0	10	4	1	0	5	12	3	0	0	8	5	15	47	27	18
Unoccupied	6	0	0	1	1	0	1	0	0	0	0	0	6	0	0	0	2	0	15	0	35	36
Core	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0

Oligomannose	5	39	2	17	6	19	80	4	4	2	25	4	3	49	48	30	19	7	5	13	7	7
Hybrid	3	11	1	16	1	2	4	1	2	1	11	8	1	4	28	18	10	10	3	0	2	0
Complex	86	49	97	66	92	79	15	95	94	97	64	88	89	47	24	50	69	82	77	87	55	57
Unoccupied	6	0	0	1	1	0	1	0	0	0	0	0	6	0	0	0	2	0	15	0	35	36
Fucose	87	14	98	32	90	69	3	79	91	96	63	73	89	50	19	32	67	36	78	86	57	43
NeuAc	26	7	43	21	35	18	0	15	31	37	4	9	33	6	6	28	13	46	18	41	14	32
Sulfation	0	0	8	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Viral derived	N17	N61	N74	N122	N149	N165	N234	N282	N331	N343	N603	N616	N657	N709	N717	N801	N1074	N1098	N1134	N1158	N1173	N1194
M9Glc	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M9	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M8	0	0	0	0	0	0	53	0	0	0	4	0	0	0	0	0	5	1	0	0	0	0
M7	0	11	0	0	0	0	15	0	1	0	9	0	0	0	22	12	10	1	0	0	0	0
M6	0	37	0	4	0	0	4	0	0	0	12	0	0	44	20	31	20	1	0	0	0	0
M5	0	12	0	0	0	0	0	0	0	0	6	0	0	16	13	0	0	1	0	0	0	0
M4	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hybrid	0	10	0	37	0	0	0	2	2	0	20	16	0	6	44	33	19	22	0	0	3	0
Fhybrid	0	0	0	8	0	0	0	0	0	13	2	0	0	0	0	0	0	2	14	0	0	0
HexNAc(3)(x)	0	0	0	7	0	0	0	2	1	0	3	0	0	0	0	0	0	3	0	0	0	0
HexNAc(3)(F)(x)	0	0	1	0	0	0	0	2	2	0	6	0	0	0	0	0	3	1	0	0	0	0
HexNAc(4)(x)	0	21	0	25	0	100	0	24	16	0	19	47	0	14	1	24	17	16	0	0	12	0
HexNAc(4)(F)(x)	34	0	35	17	57	0	0	42	47	5	19	36	0	4	0	0	20	2	86	0	22	0
HexNAc(5)(x)	8	0	0	2	0	0	0	5	7	11	0	0	0	0	0	0	6	38	0	0	18	0
HexNAc(5)(F)(x)	57	5	48	0	43	0	0	16	20	62	0	0	0	16	0	0	1	6	0	0	8	100
HexNAc(6+)(x)	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	7	0	0	0	0
HexNAc(6+)(F)(x)	0	0	16	0	0	0	0	7	2	0	0	0	0	0	0	0	0	0	0	0	38	0
Unoccupied	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Core	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Oligomannose	0	63	0	4	0	0	100	0	1	0	31	0	0	60	55	44	35	4	0	0	0	0
Hybrid	0	10	0	45	0	0	0	3	2	13	22	16	0	6	44	33	19	24	14	0	3	0
Complex	100	27	100	51	100	100	0	97	96	87	47	84	0	34	1	24	47	73	86	0	97	100
Unoccupied	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fucose	92	5	100	25	100	0	0	67	71	80	27	36	0	20	0	0	24	10	100	0	68	100
NeuAc	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0
Sulfation	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

G.3 Site-specific glycan compositions of the two RBD sites, N331 and N343

N331			
	Viral derived S	Recombinant S	Monomeric RBD
M9Glc	0%	0%	0%
M9	0%	0%	0%
M8	0%	0%	0%
M7	1%	0%	0%
M6	0%	0%	0%
M5	0%	1%	0%
M4	0%	0%	0%
M3	0%	0%	0%
FM	0%	0%	0%
Hybrid	2%	0%	0%
Fhybrid	0%	0%	0%
HexNAc(3)(x)	1%	0%	0%
HexNAc(3)(F)(x)	2%	3%	0%
HexNAc(4)(x)	16%	0%	1%
HexNAc(4)(F)(x)	47%	49%	1%
HexNAc(5)(x)	7%	0%	5%
HexNAc(5)(F)(x)	20%	40%	10%
HexNAc(6+)(x)	0%	0%	48%
HexNAc(6+)(F)(x)	2%	6%	34%
Unoccupied	0%	0%	0%
Core	1%	0%	0%

N343			
	Viral derived S	Recombinant S	Monomeric RBD
M9Glc	0%	0%	0%
M9	0%	0%	0%
M8	0%	0%	0%
M7	0%	0%	0%
M6	0%	0%	0%
M5	0%	0%	0%
M4	0%	0%	0%
M3	0%	0%	0%
FM	0%	0%	0%
Hybrid	0%	0%	0%
Fhybrid	13%	0%	0%
HexNAc(3)(x)	0%	0%	0%
HexNAc(3)(F)(x)	0%	1%	1%
HexNAc(4)(x)	0%	1%	0%
HexNAc(4)(F)(x)	5%	61%	28%
HexNAc(5)(x)	11%	0%	0%
HexNAc(5)(F)(x)	62%	32%	50%
HexNAc(6+)(x)	9%	1%	0%
HexNAc(6+)(F)(x)	0%	3%	20%
Unoccupied	0%	0%	0%
Core	0%	0%	0%

N331

Viral derived S				
Sialylated	Galactosylated	Agalactosylated	Mannose	Unoccupied
1.70%	94.46%	3.84%	0.67%	0.00%

Recombinant S				
Sialylated	Galactosylated	Agalactosylated	Mannose	Unoccupied
27.59%	39.26%	31.66%	1.37%	0.00%

Monomeric RBD				
Sialylated	Galactosylated	Agalactosylated	Mannose	Unoccupied
49.28%	47.45%	3.27%	0.47%	0.00%

N343

Viral derived S				
Sialylated	Galactosylated	Agalactosylated	Mannose	Unoccupied
0.00%	100.00%	0.00%	0.00%	0.00%

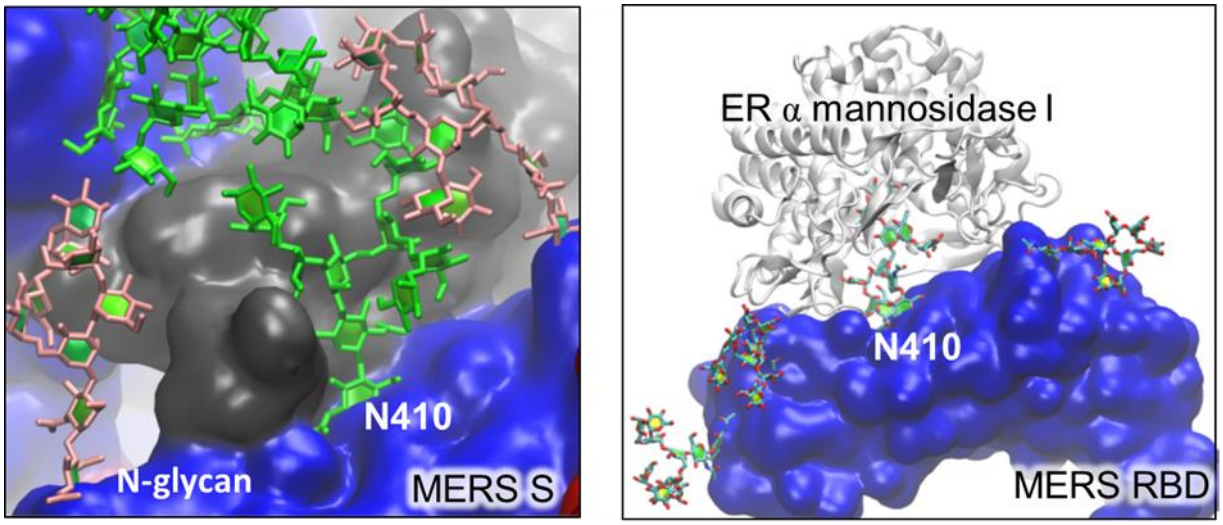
Recombinant S				
Sialylated	Galactosylated	Agalactosylated	Mannose	Unoccupied
59.09%	31.61%	9.31%	0.00%	0.00%

Monomeric RBD				
Sialylated	Galactosylated	Agalactosylated	Mannose	Unoccupied
29.77%	44.02%	26.21%	0.30%	0.00%

G.4 BLAST alignment of recombinant S proteins.

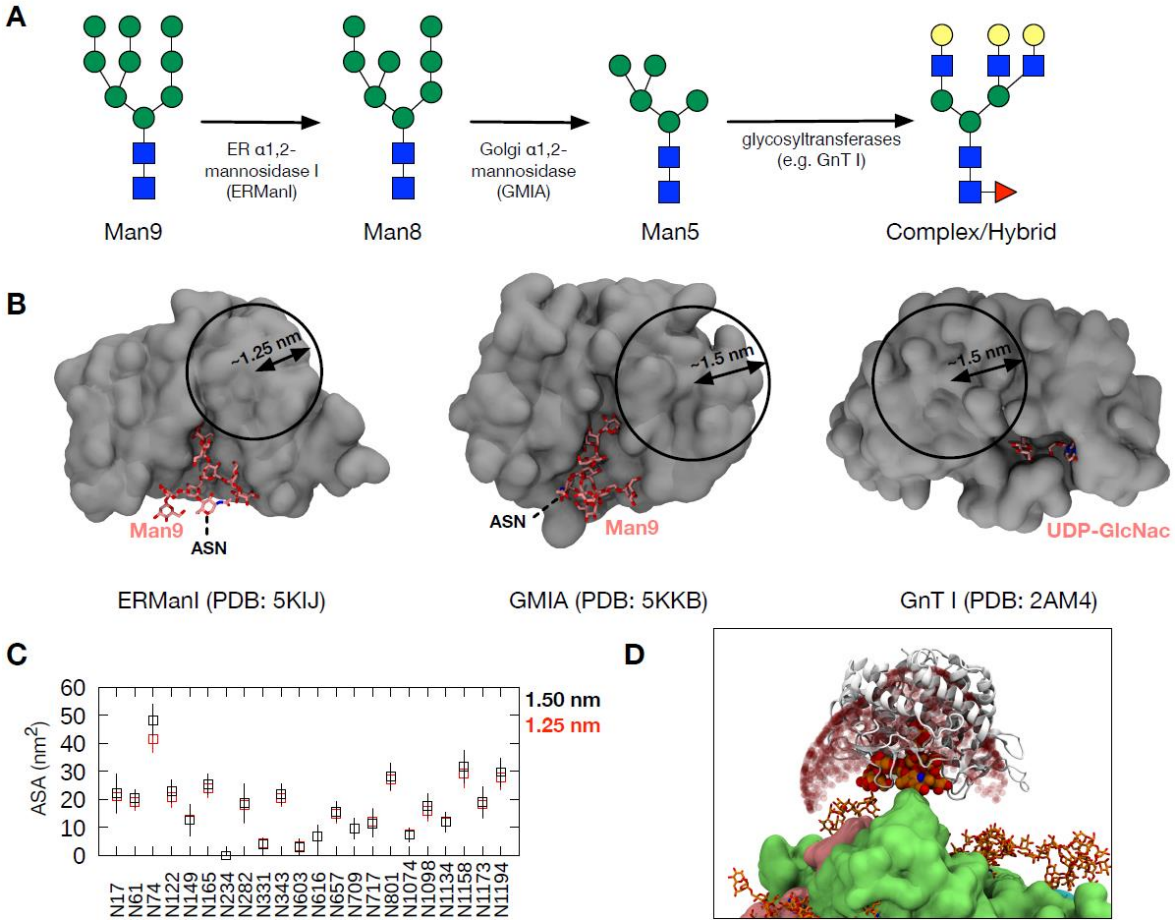
Sequences are labelled according to the principal investigator responsible for expression and purification. Protein analysed from the binding site used the McLellan construct.

G.5 Model demonstrating the differential accessibility of the N410 glycan for glycan processing enzymes between trimeric MERS S protein and monomeric MERS RBD.



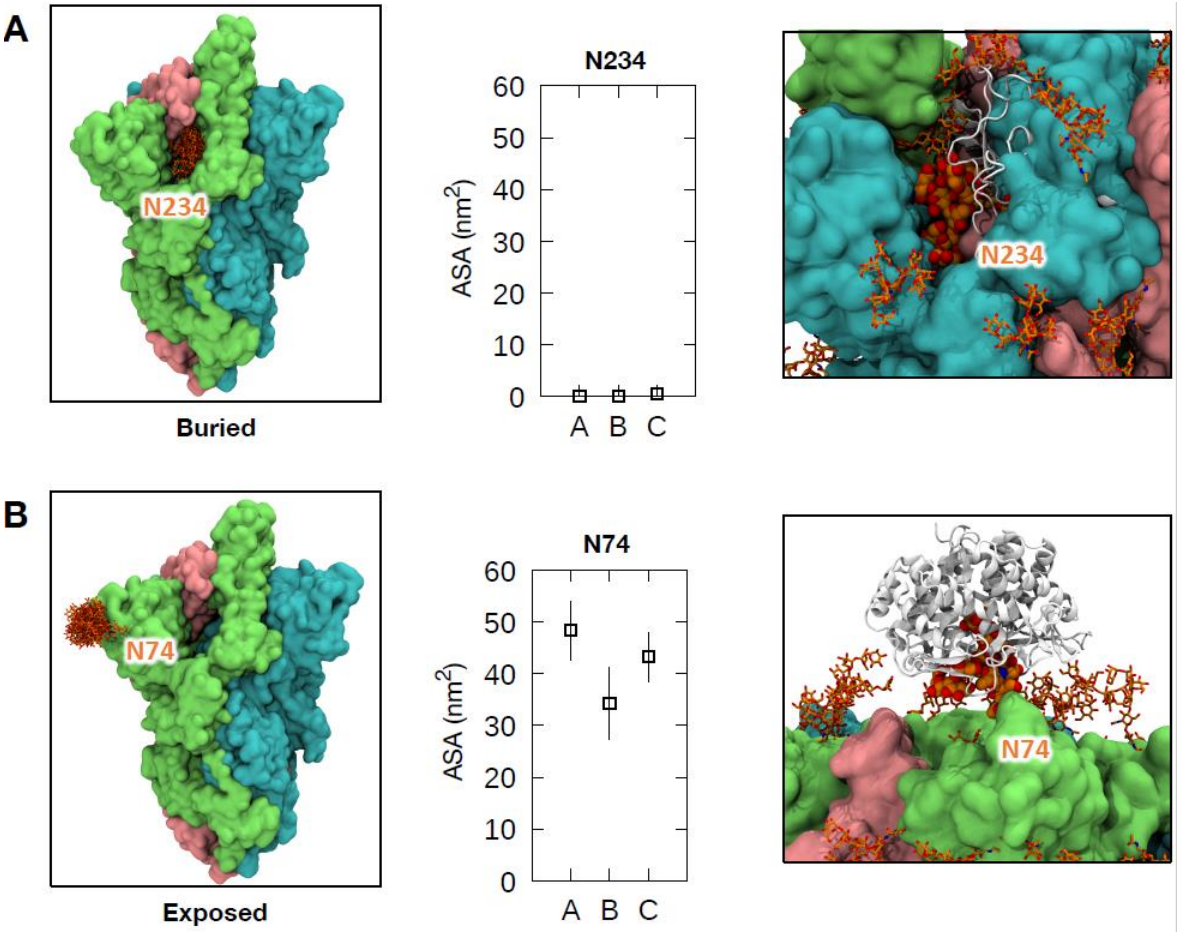
Model generated by Dr Martin Frank, Biognos AB, Sweden.

G.6 Validation of probe size for accessible surface area (ASA) measurement.



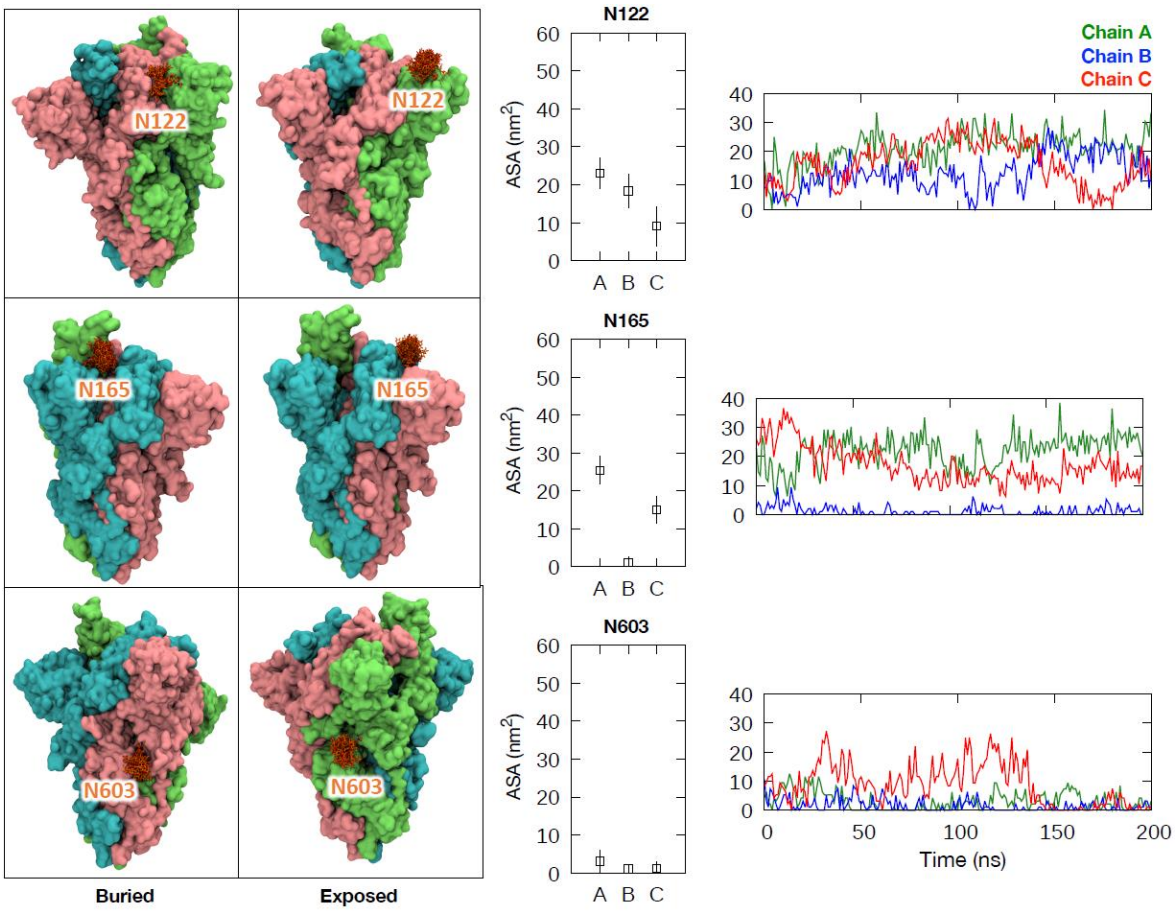
(A) A schematic of glycan processing from Man9 to complex/hybrid glycans and examples of enzymes involved. (B) Cross-section view of the crystal structures of these enzymes (grey in surface representation) bound to their substrate or substrate analogue (pink in stick representation). The circles show the approximate probe radii that could be used for glycan ASA calculation. (C) Comparison of average ASA values measured using 1.25 and 1.5 nm probe radii for Man9 in chain A of the S protein. Error bars indicate standard deviation along the simulation trajectory. (D) Accessible points (shown as red spheres) using a probe of 1.5 nm around glycan N74 (orange; van der Waals representation). S proteins are shown in surface representation in green (chain A), cyan (chain B) and pink (chain C) and other glycans shown in stick representation in orange. ERManI enzyme shown in cartoon representation in white. The simulations were performed by the laboratory of Prof. Bond, A*STAR Bioinformatics Institute, Singapore.

G.7 Examples of glycans either buried or exposed in all three chains.



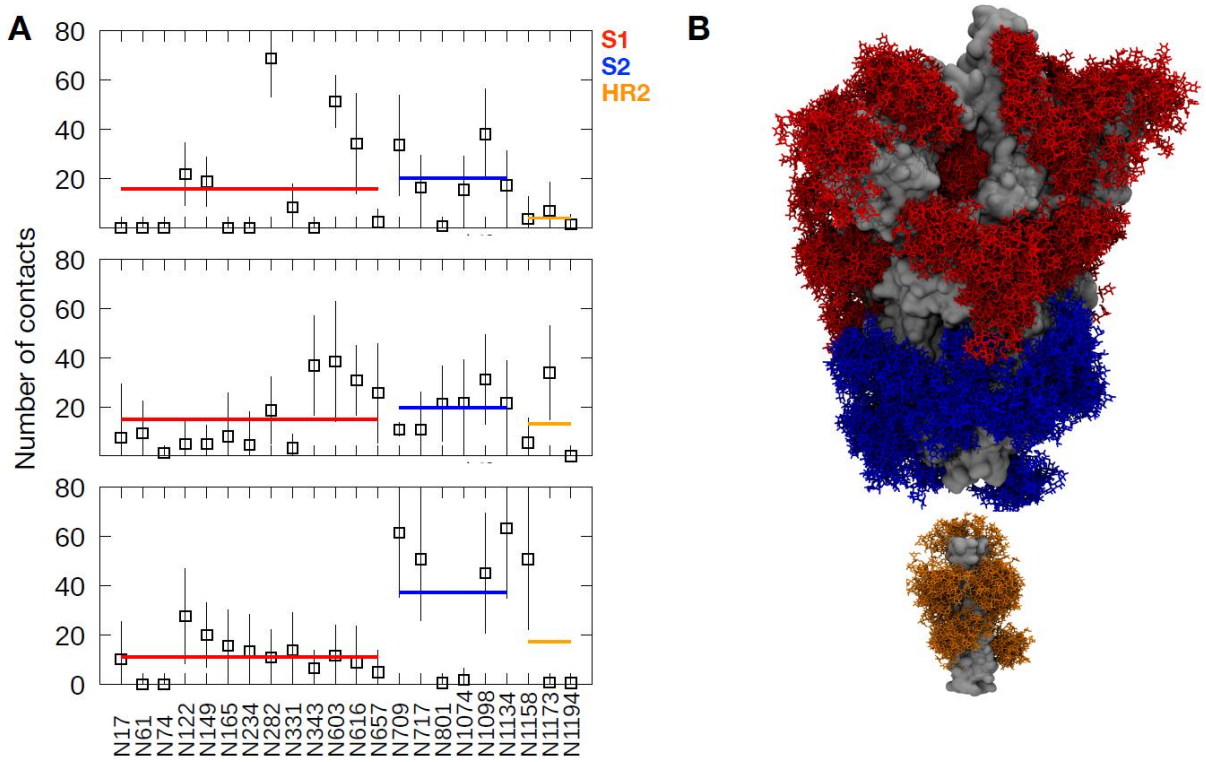
(Left) Protein shown in surface representation with the individual chains coloured differently: chain A, green; chain B, cyan; chain C, pink. The overlay of glycan snapshots taken every 5 ns for the last 50 ns of the simulation is shown in stick representation. (Middle) Average ASA values for all three chains with error bars indicating the standard deviation along the trajectory. (Right) The structure of ERManI (PDB: 5KIJ) docked onto the glycans to show overlap with the S protein in A and full accessibility of the glycan in B. ERManI is shown in white and cartoon representation. The simulations were performed by the laboratory of Prof. Bond, A*STAR Bioinformatics Institute, Singapore.

G.8 Examples of glycans with bimodal accessibility properties.



(Left) Protein is shown as in Supplementary Figure 3, with glycan snapshots taken from the portion of the simulation during which they were buried and exposed. (Middle) Average ASA values for all three chains with error bars indicating the standard deviation along the trajectory. (Right) ASA values for these glycans throughout the 200 ns simulation. The simulations were performed by the laboratory of Prof. Bond, A*STAR Bioinformatics Institute, Singapore.

G.9 Glycan-glycan contacts from MD simulation.



(A) The average number of contacts made by glycans on each glycosylation site with any other glycans on the S protein throughout the last 50 ns of the simulation. The error bars show standard deviation along the trajectory. The thick red, blue, and orange horizontal lines show the average contacts made by all glycans in the S1, S2 and HR2 regions, respectively. (B) An overlay of glycan snapshots taken every 10 ns along the 200 ns trajectory. Glycans are shown in stick representation and coloured red, blue and orange for S1, S2 and HR2 regions, respectively. S protein is shown in surface representation and coloured grey. The simulations were performed by the laboratory of Prof. Bond, A*STAR Bioinformatics Institute, Singapore.

Appendix H Subtle influence of ACE2 glycan processing on SARS-CoV-2 recognition

H.1 Site-specific glycan analysis of ACE-2

	N53	Gal	Sia	Gal no Sia	Agal
M9	0	0	0	0	0
M8	0	0	0	0	0
M7	0	0	0	0	0
M6	0	0	0	0	0
M5	0	0	0	0	0
Hybrid	0	0	0	0	0
Fhybrid	0	0	0	0	0
A1	0	0	0	0	0
FA1	2	2	0	2	0
A2/A1B	6	6	0	6	0
FA2/FA1B	23	11	8	3	13
A3/A2B	3	3	0	2	0
FA3/FA2B	43	14	0	14	29
A4/A3B	7	2	1	1	6
FA4/FA3B	16	12	4	9	4
Unoccupied	0	0	0	0	0

	N103	Gal	Sia	Gal no Sia	Agal
M9	0	0	0	0	0
M8	0	0	0	0	0
M7	0	0	0	0	0
M6	0	0	0	0	0
M5	0	0	0	0	0
Hybrid	0	0	0	0	0
Fhybrid	0	0	0	0	0
A1	3	0	0	0	3
FA1	22	2	0	2	21
A2/A1B	3	0	0	0	3
FA2/FA1B	52	45	32	13	7
A3/A2B	0	0	0	0	0
FA3/FA2B	16	11	0	11	5
A4/A3B	0	0	0	0	0
FA4/FA3B	0	0	0	0	0
Unoccupied	3	0	0	0	0

	N90	Gal	Sia	Gal no Sia	Agal
M9	0	0	0	0	0
M8	0	0	0	0	0
M7	0	0	0	0	0
M6	4	0	0	0	0
M5	0	0	0	0	0
Hybrid	0	0	0	0	0
Fhybrid	0	0	0	0	0
A1	2	1	0	1	1
FA1	0	0	0	0	0
A2/A1B	11	11	0	11	0
FA2/FA1B	52	24	5	19	27
A3/A2B	3	2	2	0	1
FA3/FA2B	27	20	10	10	7
A4/A3B	2	2	0	2	0
FA4/FA3B	0	0	0	0	0
Unoccupied	0	0	0	0	0

	N332	Gal	Sia	Gal no Sia	Agal
M9	0	0	0	0	0
M8	0	0	0	0	0
M7	0	0	0	0	0
M6	0	0	0	0	0
M5	6	0	0	0	0
Hybrid	0	0	0	0	0
Fhybrid	0	0	0	0	0
A1	0	0	0	0	0
FA1	11	0	0	0	11
A2/A1B	0	0	0	0	0
FA2/FA1B	1	1	0	1	0
A3/A2B	6	0	0	0	6
FA3/FA2B	1	1	0	1	0
A4/A3B	2	2	0	2	0
FA4/FA3B	1	1	0	0	0
Unoccupied	72	0	0	0	0

	N432	Gal	Sia	Gal no Sia	Agal
M9	0	0	0	0	0
M8	0	0	0	0	0
M7	0	0	0	0	0
M6	0	0	0	0	0
M5	1	0	0	0	0
Hybrid	0	0	0	0	0
Fhybrid	0	0	0	0	0
A1	0	0	0	0	0
FA1	0	0	0	0	0
A2/A1B	1	0	0	0	0
FA2/FA1B	10	10	4	7	0
A3/A2B	0	0	0	0	0
FA3/FA2B	25	13	0	12	12
A4/A3B	0	0	0	0	0
FA4/FA3B	0	0	0	0	0
Unoccupied	62	0	0	0	0

	N546	Gal	Sia	Gal no Sia	Agal
M9	0	0	0	0	0
M8	0	0	0	0	0
M7	0	0	0	0	0
M6	0	0	0	0	0
M5	0	0	0	0	0
Hybrid	1	0	0	0	0
Fhybrid	0	0	0	0	0
A1	0	0	0	0	0
FA1	44	4	0	4	40
A2/A1B	0	0	0	0	0
FA2/FA1B	14	14	0	14	0
A3/A2B	0	0	0	0	0
FA3/FA2B	23	0	0	0	23
A4/A3B	0	0	0	0	0
FA4/FA3B	7	0	0	0	7
Unoccupied	11	0	0	0	0

	N53	N90	N103	N332	N432	N546	Total
Oligomannose	0%	4%	0%	6%	1%	0%	2%
Sialylation	33%	17%	33%	0%	4%	0%	15%
Galactosylation	59%	60%	59%	4%	24%	18%	37%
Fucosylation	91%	79%	91%	14%	36%	88%	66%
Unoccupied	0%	0%	3%	72%	62%	11%	25%

H.2 Site-specific glycan analysis of Kif and ST6 ACE2

ACE2 +ST6

N53	Gal	Sia	Gal no Sia	Agal
M9	0	0	0	0	0
M8	0	0	0	0	0
M7	0	0	0	0	0
M6	0	0	0	0	0
M5	0	0	0	0	0
Hybrid	0	0	0	0	0
Fhybrid	0	0	0	0	0
A1	0	0	0	0	0
FA1	1	0	0	0	1
A2/A1B	7	3	3	0	4
FA2/FA1B	13	0	0	0	13
A3/A2B	0	0	0	0	0
FA3/FA2B	75	74	73	1	1
A4/A3B	0	0	0	0	0
FA4/FA3B	5	5	5	0	0
Unoccupied	0	0	0	0	0

N90	Gal	Sia	Gal no Sia	Agal
M9	0	0	0	0	0
M8	0	0	0	0	0
M7	0	0	0	0	0
M6	0	0	0	0	0
M5	0	0	0	0	0
Hybrid	0	0	0	0	0
Fhybrid	0	0	0	0	0
A1	1	1	1	0	0
FA1	1	0	0	0	1
A2/A1B	49	42	42	0	6
FA2/FA1B	36	31	31	1	4
A3/A2B	6	6	6	0	0
FA3/FA2B	8	7	7	0	1
A4/A3B	0	0	0	0	0
FA4/FA3B	0	0	0	0	0
Unoccupied	0	0	0	0	0

N103	Gal	Sia	Gal no Sia	Agal
M9	0	0	0	0	0
M8	0	0	0	0	0
M7	0	0	0	0	0
M6	0	0	0	0	0
M5	0	0	0	0	0
Hybrid	0	0	0	0	0
Fhybrid	0	0	0	0	0
A1	2	0	0	0	1
FA1	14	0	0	0	14
A2/A1B	3	0	0	0	3
FA2/FA1B	23	19	14	4	4
A3/A2B	0	0	0	0	0
FA3/FA2B	58	6	6	0	52
A4/A3B	0	0	0	0	0
FA4/FA3B	1	1	1	0	0
Unoccupied	0	0	0	0	0

N322	Gal	Sia	Gal no Sia	Agal
M9	0	0	0	0	0
M8	0	0	0	0	0
M7	0	0	0	0	0
M6	0	0	0	0	0
M5	0	0	0	0	0
Hybrid	0	0	0	0	0
Fhybrid	0	0	0	0	0
A1	0	0	0	0	0
FA1	0	0	0	0	0
A2/A1B	0	0	0	0	0
FA2/FA1B	3	0	0	0	3
A3/A2B	1	1	1	0	0
FA3/FA2B	0	0	0	0	0
A4/A3B	0	0	0	0	0
FA4/FA3B	10	10	10	0	0
Unoccupied	86	0	0	0	0

N432	Gal	Sia	Gal no Sia	Agal
M9	0	0	0	0	0
M8	0	0	0	0	0
M7	0	0	0	0	0
M6	0	0	0	0	0
M5	0	0	0	0	0
Hybrid	0	0	0	0	0
Fhybrid	0	0	0	0	0
A1	0	0	0	0	0
FA1	0	0	0	0	0
A2/A1B	0	0	0	0	0
FA2/FA1B	4	4	4	0	0
A3/A2B	0	0	0	0	0
FA3/FA2B	34	34	33	1	0
A4/A3B	0	0	0	0	0
FA4/FA3B	16	16	16	0	0
Unoccupied	47	0	0	0	0

N546	Gal	Sia	Gal no Sia	Agal
M9	0	0	0	0	0
M8	0	0	0	0	0
M7	0	0	0	0	0
M6	0	0	0	0	0
M5	19	0	0	0	0
Hybrid	2	0	2	0	0
Fhybrid	0	0	0	0	0
A1	1	1	1	0	0
FA1	0	0	0	0	0
A2/A1B	3	2	2	0	1
FA2/FA1B	32	32	32	0	0
A3/A2B	0	0	0	0	0
FA3/FA2B	37	37	37	0	0
A4/A3B	0	0	0	0	0
FA4/FA3B	0	0	0	0	0
Unoccupied	6	0	0	0	0

	N53	N90	N103	N332	N432	N546	Total
Oligomannose	0%	0%	0%	0%	0%	19%	3%
Sialylation	80%	86%	21%	12%	52%	75%	54%
Galactosylation	1%	1%	4%	0%	1%	0%	1%
Fucosylation	93%	44%	95%	13%	53%	70%	61%
Unoccupied	0%	0%	0%	86%	47%	6%	23%

ACE2 + Kif

	N53	N90	N103	N322	N432	N546
M9	100	81	79	49	50	
M8	0	1	0	0	36	
M7	0	18	20	0	0	
M6	0	0	0	0	0	
M5	0	0	0	0	0	
Hybrid	0	0	0	0	0	
Fhybrid	0	0	0	0	4	
A1	0	0	0	0	0	
FA1	0	0	0	0	0	
A2/A1B	0	0	0	0	0	
FA2/FA1B	0	0	1	0	0	
A3/A2B	0	0	0	0	0	
FA3/FA2B	0	0	0	0	0	
A4/A3B	0	0	0	0	0	
FA4/FA3B	0	0	0	0	0	
Unoccupied	0	0	0	51	10	

Total	
M9	72
M8	7
M7	8
M6	0
M5	0
Hybrid	0
Fhybrid	1
A1	0
FA1	0
A2/A1B	0
FA2/FA1B	0
A3/A2B	0
FA3/FA2B	0
A4/A3B	0
FA4/FA3B	0
Unoccupied	12

	N53	N90	N103	N322	N432	N546
Mannose	100	100	98	49	86	
Hybrid	0	0	1	0	4	
Complex	0	0	1	0	0	
Unoccupied	0	0	0	51	10	
Fucosylation	0	0	1	0	4	
Sialylation	0	0	0	0	0	

Total	
Mannose	87
Hybrid	1
Complex	0
Unoccupied	12
Fucosylation	1
Sialylation	0

H.3 Detected glycopeptides for glycosidase treated ACE2 variants

Ace2 Kif +endoH

StartAA	EndAA	Var. Pos.Protein	Sequence	Glycans	Score	Validate	XIC area
49	69	53	W.NYNTnITEENVQNMNNAGDKW.S	HexNAc(1)	205.22	True-positive	1.23E+08
80	94	90	L.AQMYPLQEIQnLTVK.L	HexNAc(1)	29.96 - 128.67	True-positive	4.39E+08
99	112	103	Q.ALQQnGSSVLSEDK.S	HexNAc(1)	29.96 - 115.30	True-positive	3.84E+07
315	328	322	F.FVSVGLPnMTQGFW.E	HexNAc(1)	29.96 - 81.43	True-positive	1.95E+10
315	328	322,323	F.FVSVGLPnmTQGFW.E	HexNAc(1)	29.96 - 69.80	True-positive	1.19E+09
542	553	546	K.CDISnSTEAGQK.L	HexNAc(1)	29.96 - 212.86	True-positive	2.16E+08

ST6 Ace2 + Sia

StartAA	EndAA	Var. Pos.Protein	Sequence	Glycans	Score	Validate	XIC area
101	112	103	L.QQnGSSVLSEDK.S	HexNAc(4)Hex(3)Fuc(1)	30	True-positive	5.25E+07
314	325		K.FFVSVGLPNMTQ.G		35.32	True-positive	7.84E+07
420	441		K.SIGLLSPDFQEDNETEINFLLK.Q		67.58 - 146.42	True-positive	3.51E+07
420	441	432	K.SIGLLSPDFQEDnETEINFLLK.Q	HexNAc(5)Hex(5)Fuc(1)	30	True-positive	2.58E+07
542	553		K.CDISNSTEAGQK.L		296.56	True-positive	4.03E+06

Ace2 + Sia

StartAA	EndAA	Var. Pos.Protein	Sequence	Glycans	Score	Validate	XIC area
75	94	90	K.EQSTLAQMYPLQEIQnLTVK.L	HexNAc(4)Hex(4)Fuc(1)	30	True-positive	2.58E+08
75	94	90	K.EQSTLAQMYPLQEIQnLTVK.L	HexNAc(5)Hex(4)Fuc(1)	29.96 - 114.06	True-positive	3.26E+08
86	94		L.QEIQNLTVK.L		105.23 - 326.29	True-positive	1.19E+08
86	94	86,90	L.qEIQnLTVK.L	HexNAc(4)Hex(4)	29.99 - 136.11	True-positive	2.42E+07
86	94	86,90	L.qEIQnLTVK.L	HexNAc(4)Hex(4)Fuc(1)	29.96 - 169.47	True-positive	4.77E+07
86	94	86,90	L.qEIQnLTVK.L	HexNAc(4)Hex(5)Fuc(1)	29.98 - 170.56	True-positive	6.44E+07
101	112	103	L.QQnGSSVLSEDK.S	HexNAc(4)Hex(3)	30	True-positive	2.94E+06
101	112		L.QQNGSSVLSEDK.S		120.03 - 215.98	True-positive	4.21E+07
101	112	101	L.qQNGSSVLSEDK.S		68.06	True-positive	3.55E+06
101	112	103	L.QQnGSSVLSEDK.S	HexNAc(3)Hex(3)	29.98 - 130.46	True-positive	3.40E+07
101	112	103	L.QQnGSSVLSEDK.S	HexNAc(3)Hex(3)Fuc(1)	29.96 - 69.91	True-positive	7.19E+07
314	325		K.FFVSVGLPNMTQ.G		95.83 - 139.82	True-positive	4.97E+08
314	325	323	K.FFVSVGLPNmTQ.G		28.00 - 45.66	True-positive	3.96E+07
314	341	332	K.FFVSVGLPNMTQGFWENSmLTDPGNVQK.A		63.59	True-positive	5.85E+06
315	325		F.FVSVGLPNMTQ.G		55.43	True-positive	1.02E+07
315	327	322	F.FVSVGLPnMTQGF.W	HexNAc(3)Hex(4)Fuc(1)	29.96 - 30.00	True-positive	8.10E+07
315	327	322	F.FVSVGLPnMTQGF.W	HexNAc(6)Hex(3)Fuc(1)	29.99 - 30.00	True-positive	1.60E+08
315	327	322	F.FVSVGLPnMTQGF.W	HexNAc(6)Hex(3)Fuc(2)	29.98 - 30.00	True-positive	7.08E+08
315	328		F.FVSVGLPNMTQGFW.E		63.41 - 93.73	True-positive	4.98E+08
316	328		F.VSVGLPNMTQGFW.E		23.61 - 163.95	True-positive	1.32E+08
420	441		K.SIGLLSPDFQEDNETEINFLLK.Q		216.8	True-positive	1.89E+09
420	441	432	K.SIGLLSPDFQEDnETEINFLLK.Q	HexNAc(3)Hex(3)Fuc(1)	30	True-positive	4.98E+07
420	441	432	K.SIGLLSPDFQEDnETEINFLLK.Q	HexNAc(4)Hex(3)	30	True-positive	5.97E+07
420	441	432	K.SIGLLSPDFQEDnETEINFLLK.Q	HexNAc(6)Hex(3)	62.17	True-positive	1.16E+07
420	441	432	K.SIGLLSPDFQEDnETEINFLLK.Q	HexNAc(4)Hex(5)Fuc(3)	29.96 - 30.00	True-positive	1.86E+07
542	553		K.CDISNSTEAGQK.L		334.12	True-positive	7.51E+07
542	553	546	K.CDISnSTEAGQK.L	HexNAc(3)Hex(3)Fuc(1)	29.96 - 585.31	True-positive	1.98E+08
542	553	546	K.CDISnSTEAGQK.L	HexNAc(3)Hex(4)Fuc(1)	381.76	True-positive	4.09E+07
542	553		K.CDISNSTEAGQK.L		41.31 - 430.43	True-positive	1.74E+08
542	553	546	K.CDISnSTEAGQK.L	HexNAc(3)Hex(3)Fuc(1)	29.96 - 30.00	True-positive	3.94E+08
542	553	546	K.CDISnSTEAGQK.L	HexNAc(5)Hex(4)Fuc(2)	29.96 - 30.00	True-positive	8.81E+06

H.4 Analysis of glycan engineered ACE2 SPR replicates

	WT	WT+Sia	ST6	ST6+Sia	+Kif	+Kif +endoH	+Fuc
Number of Repeats	30	9	12	9	6	3	3
Number of KD calculated	10	3	4	3	2	1	1
Minimum	52	48	52	45	84	58	60
Maximum	109	55	231	52	148	58	60
Range	57	7	179	7	64	0	0
Mean	76	52	116	48	116	58	60
Std. Deviation	21	4	79	4	45	0	0
Std. Error of Mean	7	2	40	2	32	0	0

	P value
Paired	
WT vs WT Sia	0.04
WT vs ST6 Sia	0.03
WT vs ST6	0.38
WT vs Kif	0.19
Unpaired	
WT vs Kif endoH	0.41
WT vs Fuc	0.46
ST6 vs ST6 Sia	0.21
ST6 vs WT Sia	0.23
Kif vs EndoH	0.48

Bibliography

1. Taubenberger, J. K. & Morens, D. M. 1918 Influenza: the Mother of All Pandemics. *Emerg. Infect. Dis.* **12**, 15–22 (2006).

2. Zheng, J. SARS-CoV-2: an Emerging Coronavirus that Causes a Global Threat. *Int. J. Biol. Sci.* **16**, 1678–1685 (2020).

3. Krammer, F. SARS-CoV-2 vaccines in development. *Nature* **586**, 516–527 (2020).

4. Sanders, R. W. & Moore, J. P. Virus vaccines: proteins prefer prolines. *Cell Host Microbe* **29**, 327–333 (2021).

5. Roser, M. & Ritchie, H. HIV / AIDS. *Our World Data* (2018).

6. Bekker, L.-G. *et al.* The complex challenges of HIV vaccine development require renewed and expanded global commitment. *Lancet* **395**, 384–388 (2020).

7. Barouch, D. H. Challenges in the development of an HIV-1 vaccine. *Nature* **455**, 613–619 (2008).

8. Gottlieb, M. S. *et al.* Pneumocystis carinii Pneumonia and Mucosal Candidiasis in Previously Healthy Homosexual Men. *N. Engl. J. Med.* **305**, 1425–1431 (1981).

9. Hymes, K. B. *et al.* Kaposi's sarcoma in homosexual men-a report of eight cases. *Lancet* **2**, 598–600 (1981).

10. U S Department of Health & Human Services. Global Statistics | HIV.gov. *2017* 1 https://www.hiv.gov/hiv-basics/overview/data-and-trends/global-statistics (2017).

11. Munis, A. M. Gene Therapy Applications of Non-Human Lentiviral Vectors. *Viruses* **12**, 1106 (2020).

12. Weiss, R. A. How does HIV cause AIDS? *Science* **260**, 1273–1279 (1993).

13. Pitisuttithum, P., Excler, J.-L. & Kim, J. Beyond RV144 Efficacy Results: An Update. *Procedia Vaccinol.* **7**, 49–56 (2013).

14. Experimental HIV Vaccine Regimen Ineffective in Preventing HIV | NIH: National Institute of Allergy and Infectious Diseases. https://www.niaid.nih.gov/news-events/experimental-hiv-vaccine-regimen-ineffective-preventing-hiv.

15. Santoro, M. M. & Perno, C. F. HIV-1 Genetic Variability and Clinical Implications. *ISRN Microbiol.* **2013**, 481314 (2013).

16. Korber, B. *et al.* Evolutionary and immunological implications of contemporary HIV-1 variation. *Br. Med. Bull.* **58**, 19–42 (2001).

17. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS pandemic. *Cold Spring Harb. Perspect. Med.* **1**, a006841–a006841 (2011).

18. Li, Y. *et al.* Eastern Chimpanzees, but Not Bonobos, Represent a Simian Immunodeficiency Virus Reservoir. *J. Virol.* **86**, 10776–10791 (2012).

19. Fischer, W. *et al.* HIV-1 and SARS-CoV-2: Patterns in the evolution of two pandemic pathogens. *Cell Host Microbe* **29**, 1093–1110 (2021).

20. Bbosa, N., Kaleebu, P. & Ssemwanga, D. HIV subtype diversity worldwide. *Current Opinion in HIV and AIDS* vol. 14 153–160 (2019).

21. Frankel, A. D. & Young, J. A. T. HIV-1: Fifteen Proteins and an RNA. *Annu. Rev. Biochem.* **67**, 1–25 (1998).

22. Scarlata, S. & Carter, C. Role of HIV-1 Gag domains in viral assembly. *Biochim. Biophys. Acta* **1614**, 62–72 (2003).

23. Philpott, S. HIV-1 Coreceptor Usage, Transmission, and Disease Progression. *Curr. HIV Res.* **1**, 217–227 (2003).

24. Gupta, R. K. *et al.* Evidence for HIV-1 cure after CCR5Δ32/Δ32 allogeneic haemopoietic stem-cell transplantation 30 months post analytical treatment interruption: a case report. *Lancet HIV* **7**, e340–e347 (2020).

25. Doms, R. W. & Moore, J. P. HIV-1 Membrane Fusion. *J. Cell Biol.* **151**, F9–F14 (2000).

26. Bebenek, K., Abbotts, J., Roberts, J. D., Wilson, S. H. & Kunkel, T. A. Specificity and Mechanism of Error-prone Replication by Human Immunodeficiency Virus-1 Reverse Transcriptase. *J. Biol. Chem.* **264**, 16948–16956 (1989).

27. Chiu, T. & Davies, D. Structure and Function of HIV-1 Integrase. *Curr. Top. Med. Chem.* **4**, 965–977 (2004).

28. Kimata, J. T., Rice, A. P. & Wang, J. Challenges and strategies for the eradication of the HIV reservoir. *Curr. Opin. Immunol.* **42**, 65–70 (2016).

29. Allen, J. D., Sanders, R. W., Doores, K. J. & Crispin, M. Harnessing post-translational modifications for next-generation HIV immunogens. *Biochem. Soc. Trans.* **46**, 691–698 (2018).

30. Seabright, G. E. The impact of sequence diversity on the glycan shield of HIV-1 immunogens. (University of Oxford, 2019).

31. Moir, S., Chun, T.-W. & Fauci, A. S. Pathogenic Mechanisms of HIV Disease. *Annu. Rev. Pathol. Mech. Dis.* **6**, 223–248 (2011).

32. Tassaneetrithep, B. *et al.* Cryptic Determinant of α4β7 Binding in the V2 Loop of HIV-1 gp120. *PLoS One* **9**, e108446 (2014).

33. Perez, L. G., Chen, H., Liao, H.-X. & Montefiori, D. C. Envelope glycoprotein binding to the integrin α4β7 is not a general property of most HIV-1 strains. *J. Virol.* **88**, 10767–77 (2014).

34. Gorny, M. K. *et al.* Functional and immunochemical cross-reactivity of V2-specific monoclonal antibodies from HIV-1-infected individuals. *Virology* **427**, 198–207 (2012).

35. Baxter, A. E. *et al.* Macrophage infection via selective capture of HIV-1-infected CD4+ T cells. *Cell Host Microbe* **16**, 711–721 (2014).

36. Battistini Garcia, S. A. & Guzman, N. *Acquired Immune Deficiency Syndrome CD4+ Count*. *StatPearls* (StatPearls Publishing, 2021).

37. Baker, S. C. Coronaviruses: From common colds to severe acute respiratory syndrome. *Pediatric Infectious Disease Journal* vol. 23 1049–1050 (2004).

38. Holmes, E. C. & Rambaut, A. Viral evolution and the emergence of SARS coronavirus. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.* **359**, 1059–1065 (2004).

39. Raj, V. S., Osterhaus, A. D. M. E., Fouchier, R. A. M. & Haagmans, B. L. MERS: emergence of a novel human coronavirus. *Curr. Opin. Virol.* **5**, 58–62 (2014).

40. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

41. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565–574 (2020).

42. Naqvi, A. A. T. *et al.* Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et Biophysica Acta - Molecular Basis of Disease* vol. 1866 165878 (2020).

43. Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **5**,

562–569 (2020).

44. Fehr, A. R. & Perlman, S. Coronaviruses: An Overview of Their Replication and Pathogenesis. in *Coronaviruses: Methods and Protocols* vol. 1282 1–23 (Springer New York, 2015).

45. Zeng, W. *et al.* Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochem. Biophys. Res. Commun.* **527**, 618–623 (2020).

46. Mollica, V., Rizzo, A. & Massari, F. The pivotal role of TMPRSS2 in coronavirus disease 2019 and prostate cancer. *Futur. Oncol.* **16**, 2029–2033 (2020).

47. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e8 (2020).

48. Naydenova, K. *et al.* Structure of the SARS-CoV-2 RNA-dependent RNA polymerase in the presence of favipiravir-RTP. *Proc. Natl. Acad. Sci.* **118**, e2021946118 (2021).

49. Robson, F. *et al.* Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. *Mol. Cell* **79**, 710–727 (2020).

50. He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).

51. Plescia, C. B. *et al.* SARS-CoV-2 viral budding and entry can be modeled using BSL-2 level virus-like particles. *Journal of Biological Chemistry* vol. 296 100103 (2021).

52. Asadi, S., Bouvier, N., Wexler, A. S. & Ristenpart, W. D. The coronavirus pandemic and aerosols: Does COVID-19 transmit via expiratory particles? *Aerosol Sci. Technol.* **54**, 635–638 (2020).

53. Sungnak, W. *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* **26**, 681–687 (2020).

54. Shah, V. K., Firmal, P., Alam, A., Ganguly, D. & Chattopadhyay, S. Overview of Immune Response During SARS-CoV-2 Infection: Lessons From the Past. *Front. Immunol.* **11**, 1949 (2020).

55. Infection control: severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): Centre for Disease Control. (2020).

56. Yi, Y., Lagniton, P. N. P., Ye, S., Li, E. & Xu, R.-H. COVID-19: what has been learned and to be learned about the novel coronavirus disease. *Int. J. Biol. Sci.* **16**, 1753–1766 (2020).

57. Turner, J. S. *et al.* SARS-CoV-2 infection induces long-lived bone marrow plasma cells in humans. *Nature* 1–5 (2021) doi:10.1038/s41586-021-03647-4.

58. Dan, J. M. *et al.* Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* **371**, (2021).

59. Pancera, M., Changela, A. & Kwong, P. D. How HIV-1 entry mechanism and broadly neutralizing antibodies guide structure-based vaccine design. *Curr. Opin. HIV AIDS* **12**, 229–240 (2017).

60. Barnes, C. O. *et al.* SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682–687 (2020).

61. Watanabe, Y., Bowden, T. A., Wilson, I. A. & Crispin, M. Exploitation of glycosylation in enveloped virus pathobiology. *Biochimica et Biophysica Acta - General Subjects* vol. 1863 1480–1497 (2019).

62. Bosch, B. J., van der Zee, R., de Haan, C. A. M. & Rottier, P. J. M. The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex. *J. Virol.* **77**, 8801–8811 (2003).

63. Harrison, S. C. Mechanism of Membrane Fusion by Viral Envelope Proteins. in *Advances in virus research* vol. 64 231–261 (Adv Virus Res, 2005).

64. Backovic, M. & Jardetzky, T. S. Class III viral membrane fusion proteins. *Curr. Opin. Struct. Biol.* **19**, 189–196 (2009).

65. He, L. *et al.* HIV-1 vaccine design through minimizing envelope metastability. *Sci. Adv.* **4**, eaau6769 (2018).

66. Sanders, R. W. & Moore, J. P. Native-like Env trimers as a platform for HIV-1 vaccine design. *Immunol. Rev.* **275**, 161–182 (2017).

67. Julien, J. P. *et al.* Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* **342**, 1477–1483 (2013).

68. McCoy, L. E. & Weiss, R. A. Neutralizing antibodies to HIV-1 induced by immunization. *J. Exp. Med.* **210**, 209–223 (2013).

69. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).

70. Barouch, D. H. *et al.* Protective efficacy of a global HIV-1 mosaic vaccine against heterologous SHIV challenges in rhesus monkeys. *Cell* **155**, 531–539 (2013).

71. Chung, A. W. *et al.* Dissecting Polyclonal Vaccine-Induced Humoral Immunity against HIV Using Systems Serology. *Cell* **163**, 988–998 (2015).

72. Rerks-Ngarm, S. *et al.* Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand. *N. Engl. J. Med.* **361**, 2209–2220 (2009).

73. Experimental HIV vaccine regimen ineffective in preventing HIV | National Institutes of Health (NIH). https://www.nih.gov/news-events/news-releases/experimental-hiv-vaccine-regimen-ineffective-preventing-hiv.

74. Sok, D. & Burton, D. R. Recent progress in broadly neutralizing antibodies to HIV. *Nature Immunology* vol. 19 1179–1188 (2018).

75. Chuang, G.-Y. *et al.* Structure-Based Design of a Soluble Prefusion-Closed HIV-1 Env Trimer with Reduced CD4 Affinity and Improved Immunogenicity. *J. Virol.* **91**, e02268-16 (2017).

76. Sanders, R. W. *et al.* A Next-Generation Cleaved, Soluble HIV-1 Env Trimer, BG505 SOSIP.664 gp140, Expresses Multiple Epitopes for Broadly Neutralizing but Not Non-Neutralizing Antibodies. *PLoS Pathog.* **9**, e1003618 (2013).

77. Sharma, S. K. *et al.* Cleavage-independent HIV-1 Env trimers engineered as soluble native spike mimetics for vaccine design. *Cell Rep.* **11**, 539–50 (2015).

78. Kulp, D. W. *et al.* Structure-based design of native-like HIV-1 envelope trimers to silence non-neutralizing epitopes and eliminate CD4 binding. *Nat. Commun.* **8**, 1655 (2017).

79. Behrens, A.-J. *et al.* Molecular Architecture of the Cleavage-Dependent Mannose Patch on a Soluble HIV-1 Envelope Glycoprotein Trimer. *J. Virol.* **91**, JVI.01894-16 (2017).

80. Stewart-Jones, G. B. E. *et al.* Trimeric HIV-1-Env Structures Define Glycan Shields from Clades A, B, and G. *Cell* **165**, 813–826 (2016).

81. Lee, J. H., de Val, N., Lyumkis, D. & Ward, A. B. Model Building and Refinement of a Natively Glycosylated HIV-1 Env Protein by High-Resolution Cryoelectron Microscopy. *Structure* **23**, 1943–1951 (2015).

82. Gristick, H. B. *et al.* Natively glycosylated HIV-1 Env structure reveals new mode for antibody recognition of the CD4-binding site. **23**, 906–915 (2016).

83. Lee, J. H., Ozorowski, G. & Ward, A. B. Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer. *Science* **351**, 1043–1048 (2016).

84. Crispin, M., Ward, A. B. & Wilson, I. A. Structure and Immune Recognition of the HIV Glycan Shield. *Annu. Rev. Biophys.* **47**, 499–523 (2018).

85. Burton, D. R. & Hangartner, L. Broadly Neutralizing Antibodies to HIV and Their Role in Vaccine Design. *Annu. Rev. Immunol.* **34**, 635–659 (2016).

86. Mascola, J. R. *et al.* Protection of macaques against vaginal transmission of a pathogenic HIV-1/SIV chimeric virus by passive infusion of neutralizing antibodies. *Nat. Med.* **6**, 207–210 (2000).

87. Hessell, A. J. *et al.* Broadly Neutralizing Monoclonal Antibodies 2F5 and 4E10 Directed against the Human Immunodeficiency Virus Type 1 gp41 Membrane-Proximal External Region Protect against Mucosal Challenge by Simian-Human Immunodeficiency Virus SHIVBa-L. *J. Virol.* **84**, 1302–1313 (2010).

88. Shingai, M. *et al.* Passive transfer of modest titers of potent and broadly neutralizing anti-HIV monoclonal antibodies block SHIV infection in macaques. *J. Exp. Med.* **211**, 2061–74 (2014).

89. Julg, B. *et al.* Protection against a mixed SHIV challenge by a broadly neutralizing antibody cocktail. *Sci. Transl. Med.* **9**, eaao4235 (2017).

90. Julg, B. *et al.* Broadly neutralizing antibodies targeting the HIV-1 envelope V2 apex confer protection against a clade C SHIV challenge. *Sci. Transl. Med.* **9**, eaal1321 (2017).

91. Shibata, R. *et al.* Neutralizing antibody directed against the HIV–1 envelope glycoprotein can completely block HIV–1/SIV chimeric virus infections of macaque monkeys. *Nat. Med.* **5**, 204–210 (1999).

92. Torrents de la Peña, A. *et al.* Improving the Immunogenicity of Native-like HIV-1 Envelope Trimers by Hyperstabilization. *Cell Rep.* **20**, 1805–1817 (2017).

93. Binley, J. M. *et al.* A Recombinant Human Immunodeficiency Virus Type 1 Envelope Glycoprotein Complex Stabilized by an Intermolecular Disulfide Bond between the gp120 and gp41 Subunits Is an Antigenic Mimic of the Trimeric Virion-Associated Structure. *J. Virol.* **74**, 627–643 (2000).

94. Stolf, B. S. *et al.* Protein disulfide isomerase and host-pathogen interaction. *ScientificWorldJournal.* **11**, 1749–1761 (2011).

95. Land, A., Zonneveld, D. & Braakman, I. Folding of HIV-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent

leader peptide cleavage. *FASEB J.* **17**, 1058–67 (2003).

96. Go, E. P., Hua, D. & Desaire, H. Glycosylation and disulfide bond analysis of transiently and stably expressed clade C HIV-1 gp140 trimers in 293T cells identifies disulfide heterogeneity present in both proteins and differences in O-linked glycosylation. *J. Proteome Res.* **13**, 4012–27 (2014).

97. Gallina, A. *et al.* Inhibitors of protein-disulfide isomerase prevent cleavage of disulfide bonds in receptor-bound glycoprotein 120 and prevent HIV-1 entry. *J. Biol. Chem.* **277**, 50579–88 (2002).

98. Go, E. P. *et al.* Native Conformation and Canonical Disulfide Bond Formation Are Interlinked Properties of HIV-1 Env Glycoproteins. *J. Virol.* **90**, 2884–2894 (2016).

99. Huang, J. *et al.* Broad and potent HIV-1 neutralization by a human antibody that binds the gp41-gp120 interface. *Nature* **515**, 138–142 (2014).

100. Behrens, A. J. *et al.* Composition and Antigenic Effects of Individual Glycan Sites of a Trimeric HIV-1 Envelope Glycoprotein. *Cell Rep.* **14**, 2695–2706 (2016).

101. Vey, M., Schäfer, W., Berghöfer, S., Klenk, H. D. & Garten, W. Maturation of the trans-Golgi network protease furin: compartmentalization of propeptide removal, substrate cleavage, and COOH-terminal truncation. *J. Cell Biol.* **127**, 1829–1842 (1994).

102. Molloy, S. S., Bresnahan, P. A., Leppla, S. H., Klimpel, K. R. & Thomas, G. Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-X-X-Arg and efficiently cleaves anthrax toxin protective antigen. *J. Biol. Chem.* **267**, 16396–16402 (1992).

103. Pritchard, L. K., Harvey, D. J., Bonomelli, C., Crispin, M. & Doores, K. J. Cell- and Protein-Directed Glycosylation of Native Cleaved HIV-1 Envelope. *J. Virol.* **89**, 8932–8944 (2015).

104. Ringe, R. P. *et al.* Cleavage strongly influences whether soluble HIV-1 envelope glycoprotein trimers adopt a native-like conformation. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18256–61 (2013).

105. Pritchard, L. K. *et al.* Structural Constraints Determine the Glycosylation of HIV-1 Envelope Trimers. *Cell Rep.* **11**, 1604–1613 (2015).

106. Walker, L. M. *et al.* Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**, 466–470 (2011).

107. Falkowska, E. *et al.* Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers. *Immunity* **40**, 657–668 (2014).

108. Binley, J. M. *et al.* Enhancing the proteolytic maturation of human immunodeficiency virus type 1 envelope glycoproteins. *J. Virol.* **76**, 2606–16 (2002).

109. Dey, A. K. *et al.* cGMP production and analysis of BG505 SOSIP.664, an extensively glycosylated, trimeric HIV-1 envelope glycoprotein vaccine candidate. *Biotechnol. Bioeng.* **115**, 885–899 (2018).

110. Dubrovskaya, V. *et al.* Targeted N-glycan deletion at the receptor-binding site retains HIV Env NFL trimer integrity and accelerates the elicited antibody response. *PLOS Pathog.* **13**, e1006614 (2017).

111. Guenaga, J. *et al.* Structure-Guided Redesign Increases the Propensity of HIV Env To Generate Highly Stable Soluble Trimers. *J. Virol.* **90**, 2806–17 (2015).

112. Mathieu, E. *et al.* A global database of COVID-19 vaccinations. *Nat. Hum. Behav.* (2021) doi:10.1038/s41562-021-01122-8.

113. Pallesen, J. *et al.* Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl. Acad. Sci.* **114**, E7348–E7357 (2017).

114. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292.e6 (2020).

115. Johnson, B. A. *et al.* Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* **591**, 293–299 (2021).

116. Hsieh, C. L. *et al.* Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **369**, 1501–1505 (2020).

117. Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3–49 (2017).

118. Behrens, A.-J. & Crispin, M. Structural principles controlling HIV envelope glycosylation. *Curr. Opin. Struct. Biol.* **44**, 125–133 (2017).

119. Li, W. *et al.* Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *Proc. Natl. Acad. Sci.* **114**, E8508–E8517 (2017).

120. Allen, J. D., Watanabe, Y., Chawla, H., Newby, M. L. & Crispin, M. Subtle Influence of ACE2 Glycan Processing on SARS-CoV-2 Recognition. *J. Mol. Biol.* **433**, 166762

(2021).

121. Higel, F. *et al.* N-glycans of complex glycosylated biopharmaceuticals and their impact on protein clearance. *Eur. J. Pharm. Biopharm.* **139**, 123–131 (2019).

122. Morniroli, D., Giannì, M. L., Consales, A., Pietrasanta, C. & Mosca, F. Human Sialome and Coronavirus Disease-2019 (COVID-19) Pandemic: An Understated Correlation? *Front. Immunol.* **11**, 1480 (2020).

123. Neelamegham, S. *et al.* Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology* **29**, 620–624 (2019).

124. Maureen E. Taylor, K. D. *Introduction to Glycobiology*.

125. Neelamegham, S. *et al.* Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology* **29**, 620–624 (2019).

126. Varki, A. *et al.* Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25**, 1323–1324 (2015).

127. Domon, B. & Costello, C. E. A Systematic Nomenclature for Carbohydrate Fragmentations in Fab-Ms Ms Spectra of Glycoconjugates. *Glycoconj. J.* **5**, 397–409 (1988).

128. Varki, A. *et al.* Glycosyltransferases. (1999).

129. Mimura, Y. *et al.* Glycosylation engineering of therapeutic IgG antibodies: challenges for the safety, functionality and efficacy. *Protein Cell* **9**, 47–62 (2018).

130. Chang, G.-D., Chen, C.-J., Lin, C.-Y., Chen, H.-C. & Chen, H. Improvement of glycosylation in insect cells with mammalian glycosyltransferases. *J. Biotechnol.* **102**, 61–71 (2003).

131. Burda, P. & Aebi, M. The dolichol pathway of N-linked glycosylation. *Biochimica et Biophysica Acta - General Subjects* vol. 1426 239–257 (1999).

132. Sanyal, S. & Menon, A. K. Specific transbilayer translocation of dolichol-linked oligosaccharides by an endoplasmic reticulum flippase. *Proc. Natl. Acad. Sci.* **106**, 767–772 (2009).

133. Shrimal, S., Cherepanova, N. A. & Gilmore, R. Cotranslational and posttranslocational N-glycosylation of proteins in the endoplasmic reticulum. *Semin. Cell Dev. Biol.* **41**, 71–78 (2015).

134. Stanley P, Taniguchi N, A. M. N-glycans. in *Essentials of Glycobiology* (ed. Varki A,

Cummings RD, Esko JD, et al.) (Cold Spring Harbor Laboratory Press, 2017). doi:10.1101/glycobiology.3e.009.

135. Hammond, C., Braakman, I. & Helenius, A. Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc. Natl. Acad. Sci.* **91**, 913–917 (1994).

136. Daniel, P. F., Winchester, B. & Warren, C. D. Mammalian alpha-mannosidases--multiple forms but a common purpose? *Glycobiology* **4**, 551–566 (1994).

137. Unligil, U. M. *et al.* X-ray crystal structure of rabbit N-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily. *EMBO J.* **19**, 5269–80 (2000).

138. Kang, M. S., Bowlin, T. L., Vijay, I. K. & Sunkara, S. P. Accumulation of pentamannose oligosaccharides in human mononuclear leukocytes by action of swainsonine, an inhibitor of glycoprotein processing. *Carbohydr. Res.* **248**, 327–337 (1993).

139. Watanabe, Y. *et al.* Signature of Antibody Domain Exchange by Native Mass Spectrometry and Collision-Induced Unfolding. *Anal. Chem.* **90**, 7325–7331 (2018).

140. Huhn, C., Selman, M. H. J., Ruhaak, L. R., Deelder, A. M. & Wuhrer, M. IgG glycosylation analysis. *Proteomics* **9**, 882–913 (2009).

141. Qasba, P. K., Ramakrishnan, B. & Boeggeman, E. Structure and function of beta -1,4-galactosyltransferase. *Curr. Drug Targets* **9**, 292–309 (2008).

142. de Vries, T., Knegtel, R. M. A., Holmes, E. H. & Macher, B. A. Fucosyltransferases: structure/function studies. *Glycobiology* **11**, 119R-128R (2001).

143. Keser, T., Pavić, T., Lauc, G. & Gornik, O. Comparison of 2-Aminobenzamide, Procainamide and RapiFluor-MS as Derivatizing Agents for High-Throughput HILIC-UPLC-FLR-MS N-glycan Analysis. *Front. Chem.* **6**, 324 (2018).

144. Rao, V., Guan, C. & Roey, P. Van. Crystal structure of endo-β-N-acetylglucosaminidase H at 1.9 å resolution: active-site geometry and substrate recognition. *Structure* **3**, 449–457 (1995).

145. Coss, K. P. *et al.* HIV-1 Glycan Density Drives the Persistence of the Mannose Patch within an Infected Individual. *J. Virol.* **90**, 11132–11144 (2016).

146. Pritchard, L. K. *et al.* Glycan clustering stabilizes the mannose patch of HIV-1 and preserves vulnerability to broadly neutralizing antibodies. *Nat. Commun.* **6**, 7479 (2015).

147. Harvey, D. J., Royle, L., Radcliffe, C. M., Rudd, P. M. & Dwek, R. A. Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry. *Anal. Biochem.* **376**, 44–60 (2008).

148. Harvey, D. J. Negative ion Mass Spectrometry for the Analysis of N -linked Glycans. *Mass Spectrom. Rev.* **39**, 586–679 (2020).

149. Ruhaak, L. R. *et al.* Glycan labeling strategies and their use in identification and quantification. *Anal. Bioanal. Chem.* **397**, 3457–3481 (2010).

150. Wilm, M. Principles of Electrospray Ionization. *Mol. Cell. Proteomics* **10**, M111.009407 (2011).

151. Cotter, R. J. Time-of-flight mass spectrometry: An increasing role in the life sciences. *Biol. Mass Spectrom.* **18**, 513–532 (1989).

152. Johnson, A. R. & Carlson, E. E. Collision-Induced Dissociation Mass Spectrometry: A Powerful Tool for Natural Product Structure Elucidation. *Anal. Chem.* **87**, 10668–10678 (2015).

153. Struwe, W. B. *et al.* Site-Specific Glycosylation of Virion-Derived HIV-1 Env Is Mimicked by a Soluble Trimeric Immunogen. *Cell Rep.* **24**, 1958-1966.e5 (2018).

154. Cipollo, J. F. & Parsons, L. M. Glycomics and glycoproteomics of viruses: Mass spectrometry applications and insights toward structure–function relationships. *Mass Spectrom. Rev.* **39**, 371–409 (2020).

155. Derking, R. *et al.* Enhancing glycan occupancy of soluble HIV-1 envelope trimers to mimic the native viral spike. *Cell Rep.* **35**, 108933 (2021).

156. Wolters, D. A., Washburn, M. P. & Yates, J. R. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Anal. Chem.* **73**, 5683–5690 (2001).

157. Link, A. J. *et al.* Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17**, 676–682 (1999).

158. Dass, C. Hyphenated Separation Techniques. in *Fundamentals of Contemporary Mass Spectrometry* 151–194 (John Wiley & Sons, Inc., 2007). doi:10.1002/9780470118498.ch5.

159. Hu, Q. *et al.* The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **40**, 430–443 (2005).

160. Olsen, J. V *et al.* Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **4**, 709–712 (2007).

161. Moini, M. Principles and Practice of Biological Mass Spectrometry. *J. Am. Chem. Soc.* **123**, 8163–8164 (2001).

162. Cao, L. *et al.* Global site-specific N-glycosylation analysis of HIV envelope glycoprotein. *Nat. Commun.* **8**, 14954 (2017).

163. Panico, M. *et al.* Mapping the complete glycoproteome of virion-derived HIV-1 gp120 provides insights into broadly neutralizing antibody binding. *Sci. Rep.* **6**, 32956 (2016).

164. Xiang, Y., Karaveg, K. & Moremen, K. W. Substrate recognition and catalysis by GH47 α-mannosidases involved in Asn-linked glycan maturation in the mammalian secretory pathway. *Proc. Natl. Acad. Sci.* **113**, E7890–E7899 (2016).

165. Behrens, A.-J. *et al.* Integrity of Glycosylation Processing of a Glycan-Depleted Trimeric HIV-1 Immunogen Targeting Key B-Cell Lineages. *J. Proteome Res.* **17**, 987–999 (2018).

166. Pollard, A. J. & Bijker, E. M. A guide to vaccinology: from basic principles to new developments. *Nat. Rev. Immunol.* **21**, 83–100 (2021).

167. Riedel, S. Edward Jenner and the History of Smallpox and Vaccination. *Baylor Univ. Med. Cent. Proc.* **18**, 21–25 (2005).

168. Greenwood, B. The contribution of vaccination to global health: past, present and future. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130433 (2014).

169. Esparza, J. A brief history of the global effort to develop a preventive HIV vaccine. *Vaccine* **31**, 3502–3518 (2013).

170. Rappuoli, R., Mandl, C. W., Black, S. & De Gregorio, E. Vaccines for the twenty-first century society. *Nat. Rev. Immunol.* **11**, 865–872 (2011).

171. Bachmann, M. F. & Jennings, G. T. Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nat. Rev. Immunol.* **10**, 787–796 (2010).

172. Zafiropoulos, A., Andersson, E., Krambovitis, E. & Borrebaeck, C. A. K. Induction of antigen-specific isotype switching by in vitro immunization of human naive B lymphocytes. *J. Immunol. Methods* **200**, 181–190 (1997).

173. Lightman, S. M., Utley, A. & Lee, K. P. Survival of Long-Lived Plasma Cells (LLPC):

Piecing Together the Puzzle. *Front. Immunol.* **10**, 965 (2019).

174. Zhang, N. & Bevan, M. J. CD8+ T Cells: Foot Soldiers of the Immune System. *Immunity* **35**, 161–168 (2011).

175. Doores, K. J. *et al.* Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 13800–5 (2010).

176. Reitter, J. N., Means, R. E. & Desrosiers, R. C. A role for carbohydrates in immune evasion in AIDS. *Nat. Med.* **4**, 679–684 (1998).

177. Scanlan, C. N. *et al.* The Broadly Neutralizing Anti-Human Immunodeficiency Virus Type 1 Antibody 2G12 Recognizes a Cluster of α1→2 Mannose Residues on the Outer Face of gp120. *J. Virol.* **76**, 7306–7321 (2002).

178. Voss, J. E. *et al.* Elicitation of Neutralizing Antibodies Targeting the V2 Apex of the HIV Envelope Trimer in a Wild-Type Animal Model. *Cell Rep.* **21**, 222–235 (2017).

179. Kong, L. *et al.* Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. *Nat. Struct. Mol. Biol.* **20**, 796–803 (2013).

180. Andrabi, R. *et al.* Glycans Function as Anchors for Antibodies and Help Drive HIV Broadly Neutralizing Antibody Development. *Immunity* **47**, 524-537.e3 (2018).

181. Pritchard, L. K. *et al.* Glycan Microheterogeneity at the PGT135 Antibody Recognition Site on HIV-1 gp120 Reveals a Molecular Mechanism for Neutralization Resistance. *J. Virol.* **89**, 6952–9 (2015).

182. Moyo, T., Kitchin, D. & Moore, P. L. Targeting the N332-supersite of the HIV-1 envelope for vaccine design. *Expert Opin. Ther. Targets* **24**, 499–509 (2020).

183. McLellan, J. S. *et al.* Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* **480**, 336–343 (2011).

184. Pancera, M. *et al.* Structural basis for diverse N-glycan recognition by HIV-1–neutralizing V1–V2–directed antibody PG16. *Nat. Struct. Mol. Biol.* **20**, 804–813 (2013).

185. Jardine, J. G. *et al.* HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen. *Science* **351**, 1458–1463 (2016).

186. Pauthner, M. *et al.* Elicitation of Robust Tier 2 Neutralizing Antibody Responses in Nonhuman Primates by HIV Envelope Trimer Immunization Using Optimized Approaches. *Immunity* **46**, 1073-1088.e6 (2017).

187. Evaluating the Safety and Immunogenicity of HIV-1 BG505 SOSIP.664 gp140 With TLR Agonist and/or Alum Adjuvants in Healthy, HIV-uninfected Adults. *Case Med. Res.* (2019) doi:10.31525/ct1-nct04177355.

188. Steichen, J. M. *et al.* HIV Vaccine Design to Target Germline Precursors of Glycan-Dependent Broadly Neutralizing Antibodies. *Immunity* **45**, 483–496 (2016).

189. Wu, X. *et al.* Rational Design of Envelope Identifies Broadly Neutralizing Human Monoclonal Antibodies to HIV-1. *Science* **329**, 856–861 (2010).

190. Jardine, J. G. *et al.* Priming a broadly neutralizing antibody response to HIV-1 using a germline-targeting immunogen. *Science* **349**, 156–161 (2015).

191. Zhou, T. *et al.* Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329**, 811–817 (2010).

192. Venkatesan, P. Preliminary phase 1 results from an HIV vaccine candidate trial. *The Lancet Microbe* **2**, e95 (2021).

193. Medina-Ramírez, M. *et al.* Design and crystal structure of a native-like HIV-1 envelope trimer that engages multiple broadly neutralizing antibody precursors in vivo. *J. Exp. Med.* **214**, 2573–2590 (2017).

194. Brouwer, P. J. M. *et al.* Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* **369**, 643–650 (2020).

195. Ju, B. *et al.* Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* **584**, 115–119 (2020).

196. Rogers, T. F. *et al.* Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science* **369**, 956–963 (2020).

197. Yuan, M. *et al.* Structural basis of a shared antibody response to SARS-CoV-2. *Science* **369**, 1119–1123 (2020).

198. Robbiani, D. F. *et al.* Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437–442 (2020).

199. Cao, Y. *et al.* Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells. *Cell* **182**, 73-84.e16 (2020).

200. Wu, Y. *et al.* A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science* **368**, 1274–1278 (2020).

201. Pinto, D. *et al.* Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290–295 (2020).

202. Seydoux, E. *et al.* Analysis of a SARS-CoV-2-Infected Individual Reveals Development of Potent Neutralizing Antibodies with Limited Somatic Mutation. *Immunity* **53**, 98-105.e5 (2020).

203. Chi, X. *et al.* A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* **369**, 650–655 (2020).

204. Shi, R. *et al.* A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature* **584**, 120–124 (2020).

205. Barnes, C. O. *et al.* Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* **182**, 828-842.e16 (2020).

206. Yu, L. & Guan, Y. Immunologic Basis for Long HCDR3s in Broadly Neutralizing Antibodies Against HIV-1. *Front. Immunol.* **5**, 250 (2014).

207. Mason, M., Sweeney, B., Cain, K., Stephens, P. & Sharfstein, S. T. Identifying bottlenecks in transient and stable production of recombinant monoclonal-antibody sequence variants in chinese hamster ovary cells. *Biotechnol. Prog.* **28**, 846–855 (2012).

208. Whitaker, N. *et al.* Developability Assessment of Physicochemical Properties and Stability Profiles of HIV-1 BG505 SOSIP.664 and BG505 SOSIP.v4.1-GT1.1 gp140 Envelope Glycoprotein Trimers as Candidate Vaccine Antigens. *J. Pharm. Sci.* **108**, 2264–2277 (2019).

209. Paragon Bioservices Selected by IAVI for Manufacture of HIV Vaccine Candidate. https://www.prnewswire.com/news-releases/paragon-bioservices-selected-by-iavi-for-manufacture-of-hiv-vaccine-candidate-300185418.html.

210. McKay, P. F. *et al.* Self-amplifying RNA SARS-CoV-2 lipid nanoparticle vaccine candidate induces high neutralizing antibody titers in mice. *Nat. Commun.* **11**, 3523 (2020).

211. Bos, R. *et al.* Ad26 vector-based COVID-19 vaccine encoding a prefusion-stabilized SARS-CoV-2 Spike immunogen induces potent humoral and cellular immune responses. *npj Vaccines* **5**, (2020).

212. Folegatti, P. M. *et al.* A Study of a Candidate COVID-19 Vaccine (COV001). *Lancet*

(2020).

213. Jackson, L. A. *et al.* An mRNA Vaccine against SARS-CoV-2 — Preliminary Report. *N. Engl. J. Med.* **383**, 1920–1931 (2020).

214. Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020).

215. Baden, L. R. *et al.* Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N. Engl. J. Med.* **384**, 403–416 (2021).

216. Ramasamy, M. N. *et al.* Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial. *Lancet* **396**, 1979–1993 (2020).

217. Corbett, K. S. *et al.* SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature* **586**, 567–571 (2020).

218. Cottrell, C. A. *et al.* Mapping the immunogenic landscape of near-native HIV-1 envelope trimers in non-human primates. *PLOS Pathog.* **16**, e1008753 (2020).

219. Sanders, R. W. *et al.* HIV-1 neutralizing antibodies induced by native-like envelope trimers. *Science* **349**, (2015).

220. Chertova, E. *et al.* Envelope Glycoprotein Incorporation, Not Shedding of Surface Envelope Glycoprotein (gp120/SU), Is the Primary Determinant of SU Content of Purified Human Immunodeficiency Virus Type 1 and Simian Immunodeficiency Virus. *J. Virol.* **76**, 5315–5325 (2002).

221. Harvey, D. J. Fragmentation of negative ions from carbohydrates: Part 3. Fragmentation of hybrid and complex N-linked glycans. *J. Am. Soc. Mass Spectrom.* **16**, 647–659 (2005).

222. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

223. Grant, T., Rohou, A. & Grigorieff, N. CisTEM, user-friendly software for single-particle image processing. *Elife* **7**, (2018).

224. Emsley, P. & Crispin, M. Structural analysis of glycoproteins: Building N-linked glycans with coot. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 256–263 (2018).

225. Aricescu, A. R., Lu, W. & Jones, E. Y. A time- and cost-efficient system for high-level

protein production in mammalian cells. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 1243–1250 (2006).

226. Huo, J. *et al.* Neutralization of SARS-CoV-2 by Destruction of the Prefusion Spike. *Cell Host Microbe* **28**, 445-454.e6 (2020).

227. Pino, P. *et al.* Trimeric SARS-CoV-2 spike proteins produced from CHO cells in bioreactors are high-quality antigens. *Processes* **8**, 1–11 (2020).

228. Yao, H. *et al.* Molecular Architecture of the SARS-CoV-2 Virus. *Cell* **183**, 730-738.e13 (2020).

229. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).

230. Hakansson-McReynolds, S., Jiang, S., Rong, L. & Caffrey, M. Solution structure of the severe acute respiratory syndrome-coronavirus heptad repeat 2 domain in the prefusion state. *J. Biol. Chem.* **281**, 11965–11971 (2006).

231. Dev, J. *et al.* Structural basis for membrane anchoring of HIV-1 envelope spike. *Science* **353**, 172–175 (2016).

232. Cai, Y. *et al.* Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, 1586–1592 (2020).

233. Eramian, D. *et al.* A composite score for predicting errors in protein structure models. *Protein Sci.* **15**, 1653–1666 (2006).

234. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* vol. 7 95–99 (1963).

235. Park, S. J. *et al.* CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **29**, 320–331 (2019).

236. Petit, C. M. *et al.* Palmitoylation of the cysteine-rich endodomain of the SARS-coronavirus spike glycoprotein is important for spike-mediated cell fusion. *Virology* **360**, 264–274 (2007).

237. Lee, J. J. *et al.* CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *J. Chem. Theory Comput.* **15**, 775–786 (2019).

238. Casares, D., Escribá, P. V. & Rosselló, C. A. Membrane lipid composition: Effect on membrane and organelle structure, function and compartmentalization and

therapeutic avenues. *Int. J. Mol. Sci.* **20**, (2019).

239. Van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: Where they are and how they behave. *Nature Reviews Molecular Cell Biology* vol. 9 112–124 (2008).

240. Van Meer, G. Lipids of the Golgi membrane. *Trends Cell Biol.* **8**, 29–33 (1998).

241. Huang, J. & Mackerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).

242. Lee, J. *et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **12**, 405–413 (2016).

243. Eslami, H. & Müller-Plathe, F. Molecular dynamics simulation in the grand canonical ensemble. *J. Comput. Chem.* **28**, 1763–1773 (2007).

244. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).

245. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

246. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

247. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).

248. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

249. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

250. Moore, P. L. *et al.* Evolution of an HIV glycan-dependent broadly neutralizing antibody epitope through immune escape. *Nat. Med.* **18**, 1688–1692 (2012).

251. Sagar, M., Wu, X., Lee, S. & Overbaugh, J. Human Immunodeficiency Virus Type 1 V1-V2 Envelope Loop Sequences Expand and Add Glycosylation Sites over the Course of Infection, and These Modifications Affect Antibody Neutralization

Sensitivity. *J. Virol.* **80**, 9586–9598 (2006).

252. Wei, X. *et al.* Antibody neutralization and escape by HIV-1. *Nature* **422**, 307–312 (2003).

253. Pegu, A., Hessell, A. J., Mascola, J. R. & Haigwood, N. L. Use of broadly neutralizing antibodies for HIV-1 prevention. *Immunological Reviews* vol. 275 296–312 (2017).

254. van Gils, M. J. & Sanders, R. W. In vivo protection by broadly neutralizing HIV antibodies. *Trends in Microbiology* vol. 22 550–551 (2014).

255. Stamatatos, L., Pancera, M. & McGuire, A. T. Germline-targeting immunogens. *Immunological Reviews* vol. 275 203–216 (2017).

256. Goo, L., Chohan, V., Nduati, R. & Overbaugh, J. Early development of broadly neutralizing antibodies in HIV-1-infected infants. *Nat. Med.* **20**, 655–658 (2014).

257. McCoy, L. E. *et al.* Incomplete Neutralization and Deviation from Sigmoidal Neutralization Curves for HIV Broadly Neutralizing Monoclonal Antibodies. *PLoS Pathog.* **11**, (2015).

258. Bonomelli, C. *et al.* The Glycan Shield of HIV Is Predominantly Oligomannose Independently of Production System or Viral Clade. *PLoS One* **6**, e23521 (2011).

259. Go, E. P. *et al.* Characterization of host-cell line specific glycosylation profiles of early transmitted/founder HIV-1 gp120 envelope proteins. *J. Proteome Res.* **12**, 1223–1234 (2013).

260. Ringe, R. P. *et al.* Influences on the Design and Purification of Soluble, Recombinant Native-Like HIV-1 Envelope Glycoprotein Trimers. *J. Virol.* **89**, 12189–12210 (2015).

261. Blattner, C. *et al.* Structural delineation of a quaternary, cleavage-dependent epitope at the gp41-gp120 interface on intact HIV-1 env trimers. *Immunity* **40**, 669–680 (2014).

262. Mouquet, H. *et al.* Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E3268–E3277 (2012).

263. Go, E. P. *et al.* Glycosylation Benchmark Profile for HIV-1 Envelope Glycoprotein Production Based on Eleven Env Trimers. *J. Virol.* **91**, 2428–2444 (2017).

264. Cao, L. *et al.* Differential processing of HIV envelope glycans on the virus and soluble recombinant trimer. *Nat. Commun.* **9**, 3693 (2018).

265. Pejchal, R. *et al.* A Potent and Broad Neutralizing Antibody Recognizes and

Penetrates the HIV Glycan Shield. *Science* **334**, 1097–1103 (2011).

266. Barnes, C. O. *et al.* Structural characterization of a highly-potent V3-glycan broadly neutralizing antibody bound to natively-glycosylated HIV-1 envelope. *Nat Commun* **9**, 1251 (2018).

267. Yolitz, J. *et al.* Signal peptide of HIV envelope protein impacts glycosylation and antigenicity of gp120. *Proc Natl Acad Sci U S A* **115**, 2443–2448 (2018).

268. Seabright, G. E. *et al.* Networks of HIV-1 Envelope Glycans Maintain Antibody Epitopes in the Face of Glycan Additions and Deletions. *Structure* **28**, 897-909.e6 (2020).

269. Lee, J. H. *et al.* A Broadly Neutralizing Antibody Targets the Dynamic HIV Envelope Trimer Apex via a Long, Rigidified, and Anionic β-Hairpin Structure. *Immunity* **46**, 690–702 (2017).

270. Shivatare, V. S. *et al.* Unprecedented Role of Hybrid N-Glycans as Ligands for HIV-1 Broadly Neutralizing Antibodies. *J Am Chem Soc* (2018) doi:10.1021/jacs.8b00896.

271. Gelderblom, H. R., Hausmann, E. H. S., Özel, M., Pauli, G. & Koch, M. A. Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. *Virology* **156**, 171–176 (1987).

272. Doms, R. W., Lamb, R. A., Rose, J. K. & Helenius, A. Folding and assembly of viral membrane proteins. *Virology* **193**, 545–562 (1993).

273. Earl, P. L., Moss, B. & Doms, R. W. Folding, interaction with GRP78-BiP, assembly, and transport of the human immunodeficiency virus type 1 envelope protein. *J. Virol.* **65**, 2047–2055 (1991).

274. Mellquist, J. L., Kasturi, L., Spitalnik, S. L. & Shakin-Eshleman, S. H. The amino acid following an Asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. *Biochemistry* **37**, 6833–6837 (1998).

275. Ruiz-Canada, C., Kelleher, D. J. & Gilmore, R. Cotranslational and Posttranslational N-Glycosylation of Polypeptides by Distinct Mammalian OST Isoforms. *Cell* **136**, 272–283 (2009).

276. Shrimal, S. & Gilmore, R. Glycosylation of closely spaced acceptor sites in human glycoproteins. *J. Cell Sci.* **126**, 5513–5523 (2013).

277. Julien, J. P. *et al.* Asymmetric recognition of the HIV-1 trimer by broadly neutralizing

antibody PG9. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 4351–4356 (2013).

278. Klasse, P. J., Ozorowski, G., Sanders, R. W. & Moore, J. P. Env Exceptionalism: Why Are HIV-1 Env Glycoproteins Atypical Immunogens? *Cell Host and Microbe* vol. 27 507–518 (2020).

279. McCoy, L. E. *et al.* Holes in the Glycan Shield of the Native HIV Envelope Are a Target of Trimer-Elicited Neutralizing Antibodies. *Cell Rep.* **16**, 2327–2338 (2016).

280. Wagh, K. *et al.* Completeness of HIV-1 Envelope Glycan Shield at Transmission Determines Neutralization Breadth. *Cell Rep.* **25**, 893-908.e7 (2018).

281. Klasse, P. J. *et al.* Epitopes for neutralizing antibodies induced by HIV-1 envelope glycoprotein BG505 SOSIP trimers in rabbits and macaques. *PLoS Pathog.* **14**, e1006913 (2018).

282. Ringe, R. P. *et al.* Closing and Opening Holes in the Glycan Shield of HIV-1 Envelope Glycoprotein SOSIP Trimers Can Redirect the Neutralizing Antibody Response to the Newly Unmasked Epitopes. *J. Virol.* **93**, 1656–1674 (2018).

283. Duan, H. *et al.* Glycan Masking Focuses Immune Responses to the HIV-1 CD4-Binding Site and Enhances Elicitation of VRC01-Class Precursor Antibodies. *Immunity* **49**, 301-311.e5 (2018).

284. Escolano, A. *et al.* Immunization expands B cells specific to HIV-1 V3 glycan in mice and macaques. *Nature* **570**, 468–473 (2019).

285. Zhou, T. *et al.* Quantification of the Impact of the HIV-1-Glycan Shield on Antibody Elicitation. *Cell Rep.* **19**, 719–732 (2017).

286. Cherepanova, N., Shrimal, S. & Gilmore, R. N-linked glycosylation and homeostasis of the endoplasmic reticulum. *Current Opinion in Cell Biology* vol. 41 57–65 (2016).

287. Kasturi, L., Eshleman, J. R., Wunner, W. H. & Shakin-Eshleman, S. H. The hydroxy amino acid in an Asn-X-Ser/Thr sequon can influence N-linked core glycosylation efficiency and the level of expression of a cell surface glycoprotein. *J. Biol. Chem.* **270**, 14756–14761 (1995).

288. Bause, E. Model studies on N glycosylation of proteins. *Biochem. Soc. Trans.* **12**, 514–517 (1984).

289. Gerber, S. *et al.* Mechanism of bacterial oligosaccharyltransferase: In vitro quantification of sequon binding and catalysis. *J. Biol. Chem.* **288**, 8849–8861 (2013).

290. Guttman, M. *et al.* CD4-induced activation in a soluble HIV-1 Env trimer. *Structure* **22**, 974–984 (2014).

291. De Taeye, S. W. *et al.* Immunogenicity of Stabilized HIV-1 Envelope Trimers with Reduced Exposure of Non-neutralizing Epitopes. *Cell* **163**, 1702–1715 (2015).

292. Pugach, P. *et al.* A native-like SOSIP.664 trimer based on an HIV-1 subtype B env gene. *J. Virol.* **89**, 3380–95 (2015).

293. Li, Y., Luo, L., Rasool, N. & Kang, C. Y. Glycosylation is necessary for the correct folding of human immunodeficiency virus gp120 in CD4 binding. *J. Virol.* **67**, 584–8 (1993).

294. Yasmeen, A. *et al.* Differential binding of neutralizing and non-neutralizing antibodies to native-like soluble HIV-1 Env trimers, uncleaved Env proteins, and monomeric subunits. *Retrovirology* **11**, 41 (2014).

295. Cupo, A. *et al.* Optimizing the production and affinity purification of HIV-1 envelope glycoprotein SOSIP trimers from transiently transfected CHO cells. *PLoS One* **14**, e0215106 (2019).

296. Dumont, J., Euwart, D., Mei, B., Estes, S. & Kshirsagar, R. Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Crit. Rev. Biotechnol.* **36**, 1110–1122 (2016).

297. Van Gils, M. J. *et al.* An HIV-1 antibody from an elite neutralizer implicates the fusion peptide as a site of vulnerability. *Nat. Microbiol.* **2**, (2016).

298. Schorcht, A. *et al.* Neutralizing Antibody Responses Induced by HIV-1 Envelope Glycoprotein SOSIP Trimers Derived from Elite Neutralizers. *J. Virol.* **94**, 1214–1234 (2020).

299. Nogal, B. *et al.* Mapping Polyclonal Antibody Responses in Non-human Primates Vaccinated with HIV Env Trimer Subunit Vaccines. *Cell Rep.* **30**, 3755-3765.e7 (2020).

300. Hu, J. K. *et al.* Murine Antibody Responses to Cleaved Soluble HIV-1 Envelope Trimers Are Highly Restricted in Specificity. *J. Virol.* **89**, 10383–10398 (2015).

301. Torrents de la Peña, A. *et al.* Similarities and differences between native HIV-1 envelope glycoprotein trimers and stabilized soluble trimer mimetics. *PLOS Pathog.* **15**, e1007920 (2019).

302. Land, A. & Braakman, I. Folding of the human immunodeficiency virus type 1

envelope glycoprotein in the endoplasmic reticulum. *Biochimie* vol. 83 783–790 (2001).

303. Snapp, E. L. *et al.* Structure and topology around the cleavage site regulate post-translational cleavage of the HIV-1 gp160 signal peptide. *Elife* **6**, (2017).

304. Sellhorn, G., Caldwell, Z., Mineart, C. & Stamatatos, L. Improving the expression of recombinant soluble HIV Envelope glycoproteins using pseudo-stable transient transfection. *Vaccine* **28**, 430–436 (2009).

305. Julien, J. P. *et al.* Design and structure of two HIV-1 clade c SOSIP.664 trimers that increase the arsenal of native-like env immunogens. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11947–11952 (2015).

306. Klasse, P. J. *et al.* Sequential and Simultaneous Immunization of Rabbits with HIV-1 Envelope Glycoprotein SOSIP.664 Trimers from Clades A, B and C. *PLOS Pathog.* **12**, e1005864 (2016).

307. Kumar, R. *et al.* Characterization of a stable HIV-1 B/C recombinant, soluble, and trimeric envelope glycoprotein (Env) highly resistant to CD4-induced conformational changes. *J. Biol. Chem.* **292**, 15849–15858 (2017).

308. Sliepen, K. *et al.* Structure and immunogenicity of a stabilized HIV-1 envelope trimer based on a group-M consensus sequence. *Nat. Commun.* **10**, (2019).

309. Torrents de la Peña, A. *et al.* Immunogenicity in rabbits of SOSIP trimers from clades A, B and C, given individually, sequentially or in combinations. *J. Virol.* JVI.01957-17 (2018) doi:10.1128/JVI.01957-17.

310. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).

311. Yang, X. *et al.* Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir. Med.* **8**, 475–481 (2020).

312. Amanat, F. & Krammer, F. SARS-CoV-2 Vaccines: Status Report. *Immunity* vol. 52 583–589 (2020).

313. Watanabe, Y. *et al.* Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat. Commun.* **11**, 2688 (2020).

314. Dalziel, M., Crispin, M., Scanlan, C. N., Zitzmann, N. & Dwek, R. A. Emerging principles for the therapeutic exploitation of glycosylation. *Science* vol. 343 (2014).

315. Scanlan, C. N., Offer, J., Zitzmann, N. & Dwek, R. A. Exploiting the defensive sugars of HIV-1 for drug and vaccine design. *Nature* vol. 446 1038–1045 (2007).

316. Walls, A. C. *et al.* Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. *Cell* **176**, 1026-1039.e15 (2019).

317. Yang, T. J. *et al.* Cryo-EM analysis of a feline coronavirus spike protein reveals a unique structure and camouflaging glycans. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 1438–1446 (2020).

318. Stertz, S. *et al.* The intracellular sites of early replication and budding of SARS-coronavirus. *Virology* **361**, 304–315 (2007).

319. Venkatagopalan, P., Daskalova, S. M., Lopez, L. A., Dolezal, K. A. & Hogue, B. G. Coronavirus envelope (E) protein remains at the site of assembly. *Virology* **478**, 75–85 (2015).

320. Ritchie, G. *et al.* Identification of N-linked carbohydrates from severe acute respiratory syndrome (SARS) spike glycoprotein. *Virology* **399**, 257–269 (2010).

321. Hargett, A. A. & Renfrow, M. B. Glycosylation of viral surface proteins probed by mass spectrometry. *Current Opinion in Virology* vol. 36 56–66 (2019).

322. Tokatlian, T. *et al.* Innate immune recognition of glycans targets HIV nanoparticle immunogens to germinal centers. *Science* **363**, 649–654 (2019).

323. Watanabe, Y. *et al.* Structure of the Lassa virus glycan shield provides a model for immunological resistance. *Proc. Natl. Acad. Sci.* **115**, 7320–7325 (2018).

324. Loke, I., Kolarich, D., Packer, N. H. & Thaysen-Andersen, M. Emerging roles of protein mannosylation in inflammation and infection. *Molecular Aspects of Medicine* vol. 51 31–55 (2016).

325. Bianchi, M. *et al.* Electron-Microscopy-Based Epitope Mapping Defines Specificities of Polyclonal Antibodies Elicited during HIV-1 BG505 Envelope Trimer Immunization. *Immunity* **49**, 288-300.e8 (2018).

326. Jardine, J. *et al.* Rational HIV immunogen design to target specific germline B cell receptors. *Science* **340**, 711–716 (2013).

327. Wei, C. J. *et al.* Cross-neutralization of 1918 and 2009 influenza viruses: Role of glycans in viral evolution and vaccine design. *Sci. Transl. Med.* **2**, 24ra21 (2010).

328. Xu, R. *et al.* Structural basis of preexisting immunity to the 2009 H1N1 pandemic

influenza virus. *Science* **328**, 357–360 (2010).

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

330. Sok, D. *et al.* Promiscuous glycan site recognition by antibodies to the high-mannose patch of gp120 broadens neutralization of HIV. *Sci. Transl. Med.* **6**, 236ra63 (2014).

331. Pinto, D. *et al.* Structural and functional analysis of a potent sarbecovirus neutralizing antibody. *bioRxiv* 2020.04.07.023903 (2020) doi:10.1101/2020.04.07.023903.

332. Allen, J. D. *et al.* Site-Specific Steric Control of SARS-CoV-2 Spike Glycosylation. *Biochemistry* acs.biochem.1c00279 (2021) doi:10.1021/acs.biochem.1c00279.

333. Liu, L. *et al.* Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* **584**, 450–456 (2020).

334. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **369**, 330–333 (2020).

335. Chen, W. H. *et al.* Yeast-expressed recombinant protein of the receptor-binding domain in SARS-CoV spike protein with deglycosylated forms as a SARS vaccine candidate. *Hum. Vaccines Immunother.* **10**, 648–658 (2014).

336. Kumar, S., Maurya, V. K., Prasad, A. K., Bhatt, M. L. B. & Saxena, S. K. Structural, glycosylation and antigenic variation between 2019 novel coronavirus (2019-nCoV) and SARS coronavirus (SARS-CoV). *VirusDisease* **31**, 13–21 (2020).

337. Carbone, F. R. & Gleeson, P. A. Carbohydrates and antigen recognition by T cells. *Glycobiology* vol. 7 725–730 (1997).

338. Chu, H. *et al.* Host and viral determinants for efficient SARS-CoV-2 infection of the human lung. *Nat. Commun.* **12**, 134 (2020).

339. Yang, Q. *et al.* Inhibition of SARS-CoV-2 viral entry upon blocking N- and O-glycan elaboration. *Elife* **9**, 1–44 (2020).

340. Zhao, P. *et al.* Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host Microbe* **28**, 586-601.e6 (2020).

341. Xu, C. & Ng, D. T. W. Glycosylation-directed quality control of protein folding. *Nature Reviews Molecular Cell Biology* vol. 16 742–752 (2015).

342. Sternberg, A. & Naujokat, C. Structural features of coronavirus SARS-CoV-2 spike

protein: Targets for vaccination. *Life Sci.* **257**, 118056 (2020).

343. Gao, Q. *et al.* Development of an inactivated vaccine candidate for SARS-CoV-2. *Science* **369**, 77–81 (2020).

344. Butler, M. Optimisation of the cellular metabolism of glycosylation for recombinant proteins produced by mammalian cell systems. *Cytotechnology* vol. 50 57–76 (2006).

345. Butler, M. & Spearman, M. The choice of mammalian cell host and possibilities for glycosylation engineering. *Current Opinion in Biotechnology* vol. 30 107–112 (2014).

346. Casalino, L. *et al.* Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **6**, 1722–1734 (2020).

347. Behrens, A. J., Struwe, W. B. & Crispin, M. Glycosylation profiling to evaluate glycoprotein immunogens against HIV-1. *Expert Review of Proteomics* vol. 14 881–890 (2017).

348. Bagdonaite, I. & Wandall, H. H. Global aspects of viral glycosylation. *Glycobiology* **28**, 443–467 (2018).

349. Wang, D. *et al.* Comprehensive Analysis of the Glycan Complement of SARS-CoV-2 Spike Proteins Using Signature Ions-Triggered Electron-Transfer/Higher-Energy Collisional Dissociation (EThcD) Mass Spectrometry. *Anal. Chem.* **92**, 14730–14739 (2020).

350. Bangaru, S. *et al.* Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate. *Science* **370**, 1089–1094 (2020).

351. Watanabe, Y. *et al.* Native-like SARS-CoV-2 Spike Glycoprotein Expressed by ChAdOx1 nCoV-19/AZD1222 Vaccine. *ACS Cent. Sci.* **7**, 594–602 (2021).

352. Gstöttner, C. *et al.* Structural and Functional Characterization of SARS-CoV-2 RBD Domains Produced in Mammalian Cells. *Anal. Chem.* **93**, 6839–6847 (2021).

353. Antonopoulos, A. *et al.* Site-specific characterization of SARS-CoV-2 spike glycoprotein receptor-binding domain. *Glycobiology* **31**, 181–187 (2021).

354. Morley, G. L. *et al.* Sensitive detection of SARS-CoV-2-Specific antibodies in dried blood spot samples. *Emerg. Infect. Dis.* **26**, 2970–2973 (2020).

355. Faustini, S. E. *et al.* Detection of antibodies to the SARS-CoV-2 spike glycoprotein in both serum and saliva enhances detection of infection. *medRxiv* (2020) doi:10.1101/2020.06.16.20133025.

356. Cook, A. M. *et al.* Validation of a combined ELISA to detect IgG, IgA and IgM antibody responses to SARS-CoV-2 in mild or moderate non-hospitalised patients. *medRxiv* (2020) doi:10.1101/2020.11.30.20229732.

357. Shields, A. *et al.* SARS-CoV-2 seroprevalence and asymptomatic viral carriage in healthcare workers: a cross-sectional study. *Thorax* **75**, 1089–1094 (2020).

358. Tai, W. *et al.* Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell. Mol. Immunol.* **17**, 613–620 (2020).

359. Stadlbauer, D. *et al.* SARS-CoV-2 Seroconversion in Humans: A Detailed Protocol for a Serological Assay, Antigen Production, and Test Setup. *Curr. Protoc. Microbiol.* **57**, (2020).

360. Johari, Y. B. *et al.* Production of trimeric SARS-CoV-2 spike protein by CHO cells for serological COVID-19 testing. *Biotechnol. Bioeng.* **118**, 1013–1021 (2021).

361. Kirchdoerfer, R. N. *et al.* Pre-fusion structure of a human coronavirus spike protein. *Nature* **531**, 118–121 (2016).

362. Bañó-Polo, M., Baldin, F., Tamborero, S., Marti-Renom, M. A. & Mingarro, I. N -glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon. *Protein Sci.* **20**, 179–186 (2011).

363. Goh, J. B. & Ng, S. K. Impact of host cell line choice on glycan profile. *Crit. Rev. Biotechnol.* **38**, 851–867 (2018).

364. Dotz, V. & Wuhrer, M. N-glycome signatures in human plasma: associations with physiology and major diseases. *FEBS Letters* vol. 593 2966–2976 (2019).

365. Boson, B. *et al.* The SARS-CoV-2 envelope and membrane proteins modulate maturation and retention of the spike protein, allowing assembly of virus-like particles. *J. Biol. Chem.* **296**, 100111 (2021).

366. Gordon, R. D. *et al.* X-ray Crystal Structures of Rabbit N-acetylglucosaminyltransferase I (GnT I) in Complex with Donor Substrate Analogues. *J. Mol. Biol.* **360**, 67–79 (2006).

367. Acharya, P. *et al.* A glycan cluster on the SARS-CoV-2 spike ectodomain is recognized by Fab-dimerized glycan-reactive antibodies. *bioRxiv* (2020) doi:10.1101/2020.06.30.178897.

368. Parodi, A. J. Protein glucosylation and its role in protein folding. *Annual Review of*

Biochemistry vol. 69 69–93 (2000).

369. Wormald, M. R. & Dwek, R. A. Glycoproteins: Glycan presentation and protein-fold stability. *Structure* vol. 7 (1999).

370. Chen, G., Zhang, Y., Trinidad, J. C. & Dann, C. Distinguishing Sulfotyrosine Containing Peptides from their Phosphotyrosine Counterparts Using Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **29**, 455–462 (2018).

371. Clarke, E. C., Nofchissey, R. A., Ye, C. & Bradfute, S. B. The iminosugars celgosivir, castanospermine and UV-4 inhibit SARS-CoV-2 replication. *Glycobiology* (2020) doi:10.1093/glycob/cwaa091.

372. Larkin, A. & Imperiali, B. The expanding horizons of asparagine-linked glycosylation. *Biochemistry* vol. 50 4411–4426 (2011).

373. Yan, R. *et al.* Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020).

374. Hayes, J. M. *et al.* Fc gamma receptor glycosylation modulates the binding of IgG glycoforms: A requirement for stable antibody interactions. *J. Proteome Res.* **13**, 5471–5485 (2014).

375. Shields, R. L. *et al.* Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human FcγRIII and antibody-dependent cellular toxicity. *J. Biol. Chem.* **277**, 26733–26740 (2002).

376. Lee, H. H. *et al.* Removal of N-Linked Glycosylation Enhances PD-L1 Detection and Predicts Anti-PD-1/PD-L1 Therapeutic Efficacy. *Cancer Cell* **36**, 168-178.e4 (2019).

377. Warren, C. J. *et al.* A glycan shield on chimpanzee CD4 protects against infection by primate lentiviruses (HIV/SIV). *Proc. Natl. Acad. Sci. U. S. A.* **166**, 11460–11469 (2019).

378. Chan, K. K. *et al.* Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science* **369**, 1261–1265 (2020).

379. Mehdipour, A. R. & Hummer, G. Dual nature of human ACE2 glycosylation in binding to SARS-CoV-2 spike. *bioRxiv* 2020.07.09.193680 (2020) doi:10.1101/2020.07.09.193680.

380. Shajahan, A. *et al.* Comprehensive characterization of N- and O- glycosylation of SARS-CoV-2 human receptor angiotensin converting enzyme 2. 2020.05.01.071688 (2020).

381. Alejandra Tortorici, M. *et al.* Structural basis for human coronavirus attachment to sialic acid receptors. *Nat. Struct. Mol. Biol.* **26**, 481–489 (2019).

382. Hulswit, R. J. G. *et al.* Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 2681–2690 (2019).

383. Baker, A. N. *et al.* The SARS-COV-2 Spike Protein Binds Sialic Acids and Enables Rapid Detection in a Lateral Flow Point of Care Diagnostic Device. *ACS Cent. Sci.* (2020) doi:10.1021/acscentsci.0c00855.

384. Engin, A. B., Engin, E. D. & Engin, A. Dual function of sialic acid in gastrointestinal SARS-CoV-2 infection. *Environmental Toxicology and Pharmacology* vol. 79 103436 (2020).

385. Clausen, T. M. *et al.* SARS-CoV-2 infection depends on cellular heparan sulfate and ACE2. *Cell* (2020) doi:10.1016/j.cell.2020.09.033.

386. Tandon, R. *et al.* Effective Inhibition of SARS-CoV-2 Entry by Heparin and Enoxaparin Derivatives. *bioRxiv* (2020) doi:10.1101/2020.06.08.140236.

387. Bermejo-Jambrina, M. *et al.* SARS-CoV-2 infection and transmission depends on heparan sulfates and is blocked by low molecular weight heparins. *bioRxiv* (2020) doi:10.1101/2020.08.18.255810.

388. Mycroft-West, C. J. *et al.* Heparin inhibits cellular invasion by SARS-CoV-2: structural dependence of the interaction of the surface protein (spike) S1 receptor binding domain with heparin. *bioRxiv* (2020) doi:10.1101/2020.04.28.066761.

389. Kim, S. Y. *et al.* Characterization of heparin and severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) spike glycoprotein binding interactions. *Antiviral Res.* **181**, 104873 (2020).

390. Ehret, J., Zimmermann, M., Eichhorn, T. & Zimmer, A. Impact of cell culture media additives on IgG glycosylation produced in Chinese hamster ovary cells. *Biotechnol. Bioeng.* **116**, 816–830 (2019).

391. Yu, X., Vasiljevic, S., Mitchell, D. A., Crispin, M. & Scanlan, C. N. Dissecting the molecular mechanism of IVIg therapy: The interaction between serum IgG and DC-SIGN is independent of antibody glycoform or Fc domain. *J. Mol. Biol.* **425**, 1253–1258 (2013).

392. Patel, K. R., Roberts, J. T., Subedi, G. P. & Barb, A. W. Restricted processing of

CD16a/Fc receptor IIIa N-glycans from primary human NK cells impacts structure and function. *J. Biol. Chem.* **293**, 3477–3489 (2018).

393. Chang, V. T. *et al.* Glycoprotein Structural Genomics: Solving the Glycosylation Problem. *Structure* **15**, 267–273 (2007).

394. Milanetti, E. *et al.* In-Silico evidence for two receptors based strategy of SARS-CoV-2. *bioRxiv* (2020) doi:10.1101/2020.03.24.006197.

395. Monteil, V. *et al.* Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. *Cell* **181**, 905-913.e7 (2020).

396. Miao, X. *et al.* A novel biparatopic hybrid antibody-ACE2 fusion that blocks SARS-CoV-2 infection: implications for therapy. *MAbs* **12**, (2020).

397. Pang, X., Cui, Y. & Zhu, Y. Recombinant human ACE2: potential therapeutics of SARS-CoV-2 infection and its complication. *Acta Pharmacologica Sinica* vol. 41 1255–1257 (2020).

398. Zhang, H., Penninger, J. M., Li, Y., Zhong, N. & Slutsky, A. S. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Med.* **46**, 586–590 (2020).

399. Brouwer, P. J. M. M. *et al.* Immunofocusing and enhancing autologous Tier-2 HIV-1 neutralization by displaying Env trimers on two-component protein nanoparticles. *npj Vaccines* **6**, 24 (2021).

400. Sicari, D., Chatziioannou, A., Koutsandreas, T., Sitia, R. & Chevet, E. Role of the early secretory pathway in SARS-CoV-2 infection. *J. Cell Biol.* **219**, (2020).

401. Mauro, V. P. & Chappell, S. A. A critical analysis of codon optimization in human therapeutics. *Trends Mol. Med.* **20**, 604–613 (2014).

402. Kong, R. *et al.* Fusion peptide of HIV-1 as a site of vulnerability to neutralizing antibody. *Science* **352**, 828–833 (2016).

403. Xu, K. *et al.* Epitope-based vaccine design yields fusion peptide-directed antibodies that neutralize diverse strains of HIV-1. *Nat. Med.* **24**, 857–867 (2018).

404. Lemmin, T., Soto, C., Stuckey, J. & Kwong, P. D. Microsecond Dynamics and Network Analysis of the HIV-1 SOSIP Env Trimer Reveal Collective Behavior and Conserved Microdomains of the Glycan Shield. *Structure* **25**, 1631-1639.e2 (2017).

405. Lee, C.-C. D. *et al.* A cross-neutralizing antibody between HIV-1 and influenza virus.

PLoS Pathog. **17**, e1009407 (2021).

406. Andrabi, R. *et al.* The Chimpanzee SIV Envelope Trimer: Structure and Deployment as an HIV Vaccine Template. *Cell Rep.* **27**, 2426-2441.e6 (2019).

407. Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **19**, 409–424 (2021).

408. Li, Z. *et al.* The human coronavirus HCoV-229E S-protein structure and receptor binding. *Elife* **8**, (2019).

409. Walls, A. C. *et al.* Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* **23**, 899–905 (2016).

410. Abdul-Rasool, S. & Fielding, B. C. Understanding Human Coronavirus HCoV-NL63. *Open Virol. J.* **4**, 76–84 (2010).

411. Williams, W. B. *et al.* Fab-dimerized glycan-reactive antibodies are a structural category of natural antibodies. *Cell* **184**, 2955-2972.e25 (2021).

412. Heath, P. T. *et al.* Safety and Efficacy of NVX-CoV2373 Covid-19 Vaccine. *N. Engl. J. Med.* NEJMoa2107659 (2021) doi:10.1056/nejmoa2107659.

413. Liao, H.-Y. *et al.* Impact of Glycosylation on SARS-CoV-2 Infection and Broadly Protective Vaccine Design. *bioRxiv* 2021.05.25.445523 (2021) doi:10.1101/2021.05.25.445523.

414. Seabright, G. E., Doores, K. J., Burton, D. R. & Crispin, M. Protein and Glycan Mimicry in HIV Vaccine Design. *J. Mol. Biol.* **431**, 2223–2247 (2019).

415. SYNAPT G2-Si brochure | Waters. https://www.waters.com/waters/library.htm?locale=en_US&cid=134740622&lid=134743723.

416. Planet Orbitrap- Orbitrap Fusion Mass Spectrometer Schematic. https://planetorbitrap.com/orbitrap-fusion#tab:schematic.

List of Publications

1. Allen, J. D. et al. Site-Specific Steric Control of SARS-CoV-2 Spike Glycosylation. Biochemistry 60, 2153–2169 (2021).

2. Allen, J. D., Watanabe, Y., Chawla, H., Newby, M. L. & Crispin, M. Subtle Influence of ACE2 Glycan Processing on SARS-CoV-2 Recognition. J. Mol. Biol. 433, 166762 (2021).

3. Antanasijevic, A. et al. Polyclonal antibody responses to HIV Env immunogens resolved using cryoEM. bioRxiv (2021).

4. Brouwer, P. J. M. et al. Immunofocusing and enhancing autologous Tier-2 HIV-1 neutralization by displaying Env trimers on two-component protein nanoparticles. npj Vaccines 6, 24 (2021).

5. Brouwer, P. J. M. et al. Two-component spike nanoparticle vaccine protects macaques from SARS-CoV-2 infection. Cell (2021) doi:10.1016/j.cell.2021.01.035.

6. Cook, A. M. et al. Validation of a combined ELISA to detect IgG, IgA and IgM antibody responses to SARS-CoV-2 in mild or moderate non-hospitalised patients. J. Immunol. Methods 494, 113046 (2021).

7. Derking, R. et al. Enhancing glycan occupancy of soluble HIV-1 envelope trimers to mimic the native viral spike. Cell Rep. 35, 108933 (2021).

8. Faustini, S. E. et al. Development of a high sensitivity ELISA detecting IgG, A & M antibodies to the SARS-CoV-2 spike glycoprotein in serum and saliva. Immunology 00, 1–13 (2021).

9. Hangartner, L. et al. Effector function does not contribute to protection from virus challenge by a highly potent HIV broadly neutralizing antibody in nonhuman primates. Sci. Transl. Med. 13, (2021).

10. Khan, A. O. et al. Stimulation of vascular organoids with SARS-CoV-2 antigens increases endothelial permeability and regulates vasculopathy. medRxiv (2021).

11. Lee, C.-C. D. et al. A cross-neutralizing antibody between HIV-1 and influenza virus. PLoS Pathog. 17, e1009407 (2021).

12. Margolin, E. A. et al. Site-specific glycosylation of recombinant viral glycoproteins produced in Nicotiana benthamiana. Front. Plant Sci. 12, 1473 (2021).

13. Perez-Toledo, M. et al. SARS-CoV-2-specific IgG1/IgG3 but not IgM in children with Pediatric Inflammatory Multi-System Syndrome. Pediatr. allergy Immunol. (2021) doi:10.1111/pai.13504.

14. Sanchez, I. D. M. et al. Novel trimer-only (TO) producing HIV-1 envelope glycoprotein constructs for inducing broadly neutralizing antibody responses by genetic vaccination. J. Int. AIDS Soc. 24, 37–39 (2021).

15. Antanasijevic, A. et al. Structural and functional evaluation of de novo-designed, two-component nanoparticle carriers for HIV Env trimer immunogens. PLoS Pathog. 16, (2020).

16. Kumar, S. et al. Neutralizing antibodies induced by first-generation gp41-stabilized HIV-1 envelope trimers and nanoparticles. bioRxiv (2020) doi:10.1101/2020.12.02.408328.

17. Schorcht, A. et al. Neutralizing Antibody Responses Induced by HIV-1 Envelope Glycoprotein SOSIP Trimers Derived from Elite Neutralizers. J. Virol. 94, 1214–1234 (2020).

18. Seabright, G. E. G. E. et al. Networks of HIV-1 Envelope Glycans Maintain Antibody Epitopes in the Face of Glycan Additions and Deletions. Structure 28, 897-909.e6 (2020).

19. Shields, A. et al. SARS-CoV-2 seroprevalence and asymptomatic viral carriage in healthcare workers: A cross-sectional study. Thorax 75, 1089–1094 (2020).

20. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. Science 369, 330–333 (2020).

21. Watanabe, Y. et al. Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. Nat. Commun. 11, 2688 (2020).

22. Andrabi, R. et al. The Chimpanzee SIV Envelope Trimer: Structure and Deployment as an HIV Vaccine Template. Cell Rep. 27, 2426-2441.e6 (2019).

23. Brouwer, P. J. M. et al. Enhancing and shaping the immunogenicity of native-like HIV-1 envelope trimers with a two-component protein nanoparticle. Nat. Commun. 10, (2019).

24. Tokatlian, T. et al. Innate immune recognition of glycans targets HIV nanoparticle immunogens to germinal centers. Science 363, eaat9120 (2019).

25. Torrents de la Peña, A. et al. Similarities and differences between native HIV-1 envelope glycoprotein trimers and stabilized soluble trimer mimetics. PLoS Pathog. 15, (2019).

26. Aldon, Y. et al. Rational Design of DNA-Expressed Stabilized Native-Like HIV-1 Envelope Trimers. Cell Rep. 24, 3324-3338.e5 (2018).

27. Allen, J. D., Sanders, R. W., Doores, K. J. & Crispin, M. Harnessing post-translational modifications for next-generation HIV immunogens. Biochem. Soc. Trans. 46, 691–698 (2018).

28. Harvey, D. J. et al. Collision Cross Sections and Ion Mobility Separation of Fragment Ions from Complex N-Glycans. J. Am. Soc. Mass Spectrom. 29, 1250–1261 (2018).

29. He, L. et al. HIV-1 vaccine design through minimizing envelope metastability. Sci. Adv. 4, eaau6769 (2018).

30. Struwe, W. B. et al. Site-Specific Glycosylation of Virion-Derived HIV-1 Env Is Mimicked by a Soluble Trimeric Immunogen. Cell Rep. 24, 1958-1966.e5 (2018).

31. Watanabe, Y. et al. Structure of the Lassa virus glycan shield provides a model for immunological resistance. Proc. Natl. Acad. Sci. U. S. A. 115, 7320–7325 (2018).

32. Watanabe, Y. et al. Signature of Antibody Domain Exchange by Native Mass Spectrometry and Collision-Induced Unfolding. Anal. Chem. 90, 7325–7331 (2018).

33. Young, G. et al. Quantitative mass imaging of single biological macromolecules. Science 360, 423–427 (2018).

34. Blundell, P. A. P. A., Le, N. P. L. N. P. L., Allen, J., Watanabe, Y. & Pleass, R. J. R. J. Engineering the fragment crystallizable (Fc) region of human IgG1 multimers and monomers to fine-tune interactions with sialic acid-dependent receptors. J. Biol. Chem. 292, 12994–13007 (2017).