Protein and glycan mimicry in HIV vaccine design

Gemma E. Seabright\textsuperscript{a,b}, Katie J. Doores\textsuperscript{c}, Dennis R. Burton\textsuperscript{d} and Max Crispin\textsuperscript{b,d}

\textbf{a.} Oxford Glycobiology Institute, Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, UK

\textbf{b.} School of Biological Sciences, University of Southampton, Southampton, SO17 1BJ, UK

\textbf{c.} Department of Infectious Diseases, King's College London, Guy's Hospital, London, SE1 9RT, UK

\textbf{d.} Department of Immunology and Microbiology, the Scripps Centre for HIV/AIDS Vaccine Immunology and Immunogen Discovery (CHAVI-ID), International AIDS Vaccine Initiative Neutralizing Antibody Centre, Scripps Research, La Jolla, CA, USA

*Corresponding author

Email: max.crispin@soton.ac.uk

Phone: +44 (0) 2380 594268

Postal address: Building 85, Biological Sciences, University of Southampton, Southampton, SO17 1BJ, UK
Abstract

Antigenic mimicry is a fundamental tenet of structure-based vaccinology. Vaccine strategies for the human immunodeficiency virus type 1 (HIV-1) focus on the mimicry of its envelope spike (Env) due to its exposed location on the viral membrane and role in mediating infection. However, the virus has evolved to minimise the immunogenicity of conserved epitopes on the envelope spike. This principle is starkly illustrated by the presence of an extensive array of host-derived glycans which act to shield the underlying protein from antibody recognition. Despite these hurdles, a subset of HIV-infected individuals eventually develop broadly neutralising antibodies that recognise these virally-presented glycans. Effective HIV-1 immunogens are therefore likely to involve some degree of mimicry of both the protein and glycan components of Env. As such, considerable efforts have been made to characterise the structure of the envelope spike and its glycan shield. This review summarises the recent progress made in this field, with an emphasis on our growing understanding of the factors shaping the glycan shield of Env derived from both virus and soluble immunogens. We argue that recombinant mimics of the envelope spike are currently capable of capturing many features of the native viral glycan shield. Finally, we explore strategies through which the immunogenicity of Env glycans may be enhanced in the development of future immunogens.

Keywords: human immunodeficiency virus, vaccinology, antibodies, glycosylation, structure

Abbreviations/Glossary: HIV-1, human immunodeficiency virus type 1; Env, envelope spike; AIDS, acquired immune deficiency syndrome; bnAb(s), broadly neutralising antibody(ies); nAb(s), neutralising antibody(ies); gl-bnAb(s), germline-bnAb(s); CD4bs, CD4 binding site; CCR5, C-C chemokine receptor type 5; CXCR4, C-X-C chemokine receptor type 4; TF, transmitted/founder; EM, electron microscopy; PNGS, potential N-glycosylation sites; BCR, B cell receptor, IMP, intrinsic mannose patch; TAMP, trimer associated mannose patch; HCDR3, third heavy chain complementarity-determining regions; CDR, complementarity-determining regions; LOS, lipooligosaccharides; SP, signal peptide; MPER, membrane proximal external region; TM, transmembrane region; CT, cytoplasmic tail; HR1/2, heptad repeat 1 or 2; NFL, native flexibly linked; SC, single-chain; UFO, uncleaved prefusion-optimised; PBMC, peripheral blood mononuclear cell; PNS, peripheral nervous system; CHO, Chinese hamster ovary; HEK, human embryonic kidney; cGMP, current good manufacturing practices; Glc, glucose; Man, mannose; GlcNac, N-acetylglicosamine; Gal, galactose; Fuc, fucose; Neu5Ac, N-acetyleneuraminic acid (sialic acid); GlcN, glucosamine; KDO, 2-keto-3-deoxy-d-manno-octulosonic acid; ER, endoplasmic reticulum; α-man I and II, α-mannosidase I and II; GnT I, N-acetylglicosaminytransferase I.
1. Challenges facing HIV-1 vaccine design

Vaccines typically contain or mimic parts or all of a pathogen, such as an attenuated strain or recombinant soluble surface protein, to prime the immune system to produce an effective response upon future exposure to that pathogen. This strategy has proved to be very successful in the past, famously resulting in the complete eradication of the smallpox virus [1], and more recently in a protective vaccine against Ebola virus [2]. Despite significant efforts, a vaccine capable of eliciting a protective response against the human immunodeficiency virus type 1 (HIV-1) has proved elusive [3].

Both antibodies and cytotoxic T lymphocytes are produced upon infection with HIV-1. However, the virus has evolved several features that undermine immunological control and eradication of infection, most notably, very high antigenic diversity and the establishment of a latent viral reservoir. While treatment with antiretroviral drugs can extend the life expectancy of infected individuals to near-normal [4, 5], drug-resistance has been documented for every class of antiretroviral currently in use [6], and treatment regimens are often accompanied by adverse side-effects and low levels of adherence. Furthermore, cessation of therapy results in rapid viral rebound [7]. If left untreated, HIV-1 infection results in diminished numbers of CD4+ T cells (the major viral target cell), causing acquired immune deficiency syndrome (AIDS) and death. While HIV-1 cure strategies are an important and viable field of research [8], the development of an effective prophylactic vaccine remains a primary goal in the effort to control the HIV-1 pandemic.

Analysis of the immune response of infected individuals has renewed optimism that a vaccine may be a tractable goal [9, 10]. A subset of HIV-1 infected patients are able to generate antibodies of sufficient breadth and potency to neutralise the vast majority of circulating HIV-1 isolates [11-13]. Although these broadly neutralising antibodies (bnAbs) are unable to clear the virus from the infected individual, they are able to protect non-human primates [14-25] and humanised mice [26-32] from viral challenge when passively administered. Importantly, these antibodies are protective at concentrations achievable by vaccination in other settings [15, 17]. Taken together, these observations provide some support for the hypothesis that a vaccine can be developed capable of generating a protective antibody response against HIV-1.

All known bnAbs are directed against the envelope spike (Env) [33-35], the only viral protein on the virus surface (Fig. 1A). Therefore, while the contribution of T cells in the development of an antibody response is critical [36-38], considerable research efforts have been directed at the development of stable, recombinant mimics of the envelope spike in order to elicit a B cell response [39]. Central to this strategy is the hypothesis that antigenic mimicry of a vaccine candidate is essential for the induction of an antibody response against that antigen [37, 40]. However, HIV-1 has evolved
under immense selection pressure by the humoral immune system and consequently many of the most valuable bnAb epitopes are inherently poor immunogens. One manifestation of this is an extensive array of host-derived N-glycans which surrounds the envelope spike to create a largely immunologically ‘self’ glycan shield (Fig. 1B). While originally thought to protect the underlying protein surface from immune recognition, the discovery that many bnAbs can develop that recognise glycan epitopes has exposed the glycan shield itself as a potential target for vaccine design [41]. Scanlan et al. previously highlighted the apparent contradiction in that HIV-1 glycans have evolved as an adaptation for virus survival and yet have emerged as targets for vaccine design [42]. It is therefore important that Env-based immunogens are able to mimic effectively both the protein and glycan components of the envelope spike, though strategies that tackle the poor immunogenicity of the glycan epitopes are likely to be required.

In order to inform the rational design of Env immunogens a detailed understanding of both the composition of the glycan shield, and the structural rules governing its formation are required. This review will discuss recent contributions to the field of HIV-1 vaccine design, specifically the principles governing HIV-1 glycosylation and how this can be used to help select candidate immunogens. We also discuss the strategies being explored with the aim of boosting the immunogenicity of Env-based vaccines.
Figure 1. Structure of the HIV-1 virion and the envelope spike. (A) Graphic depicting the structure of the HIV-1 virion. Approximately 14 envelope spikes are displayed on the surface of the virion (mean for one HIV-1 isolate), embedded into the host cell-derived lipid membrane [43]. (B) Model of a fully glycosylated envelope spike (glycans in cyan sticks) based on PDB: 5ACO [44]. Glycans were added according to Behrens et al., 2016 [45]. The envelope spike is a trimer of non-covalently associated gp120 (light grey) and gp41 (dark grey) heterodimers. The gp120 subunits contain the CD4 receptor and CCR5 or CXCR4 co-receptor binding sites. Upon binding, the trimer undergoes substantial conformational changes that enable the gp41 subunits to drive fusion of the viral and host cell membranes. The membrane proximal external region (MPER), transmembrane domain (TM) and cytoplasmic tail (CT) are not present on the structure and are shown for one protomer in cartoon along with the lipid membrane for orientation. (C) Left: Schematic representation of the primary structure of Env (top) and the soluble immunogen, BG505 SOSIP.664 (bottom). Variable regions (V1-5) are shown in colour, constant regions (C1-5) are shown in light grey, and gp41 is shown in dark grey. The envelope spike has approximately 25 potential N-glycosylation sites per gp120, and 4 per gp41 (cyan forks; mean across many isolates) [46]. SOSIP.664 modifications are annotated in magenta, with optional purification tag coloured green. FP = fusion peptide, HR1/2 = heptad repeat 1 and 2. Right: Model of a de-glycosylated envelope spike (as in B), with variable loops coloured accordingly.

1.1 The antibody response

Neutralising antibodies (nAbs) typically work by binding an antigen on the viral surface to prevent the virus from infecting the host cell, and correlate with protection in the majority of licensed vaccines [47]. As the only viral protein on the virion surface, and playing a key role in mediating infection, the envelope spike is the sole target for nAbs. However, several fundamental features of HIV-1 biology hinder the development of nAbs in both a vaccine and infection setting. The functional envelope spike is a trimer of non-covalently associated gp120-gp41 heterodimers (Fig. 1), generated by furin cleavage of a gp160 precursor polypeptide. Host cell tropism and attachment is mediated by the gp120 subunits, which contain the CD4 receptor and co-receptor (CCR5 or CXCR4) binding sites. Once bound, substantial conformational changes enable the gp41 subunits to drive fusion of the viral and host cell membranes [48]. Antibodies capable of binding the spike in its functional conformation prevent this occurring. However, functional Envs are few and far between, with only approximately 14 spikes per viral particle (mean for one HIV-1 isolate) [43]. The wide spacing of the envelope spikes is disadvantageous for the host antibody response as B cells are more effectively activated by repetitive and organised structures [49].

The virus also produces an abundance of non-functional envelope spikes in the form of uncleaved gp160, and non-functional monomeric subunits such as soluble gp120 and gp41 stumps. These can arise either through the improper processing of the spike, or its later disintegration (Fig. 1A). This ‘viral debris’ displays immunodominant epitopes that are either occluded or absent on the functional trimer (e.g. the inner surface of gp120), and acts to divert the host antibody response [50-
Thus, the initial antibody response, arising over the first few weeks of infection, is incapable of binding the functional envelope spike and is ‘non-neutralising’ [56].

A further nAb evasion feature of HIV-1 is the relatively poor accessibility of its conserved epitopes. The CD4 binding site (CD4bs), for example, is a highly conserved region essential for infectivity and thus represents a potentially vulnerable site for antibody neutralisation. However, its recessed location within the trimer interface, surrounded by N-glycans, reduces the accessibility of this valuable collection of epitopes for nAb recognition. Interestingly, llamas and cows are capable of generating nAbs against the CD4bs region following immunisation with Env [57, 58]. Llamas naturally produce heavy-chain only antibodies that are much smaller than conventional antibodies, while cow antibodies contain very long third heavy chain complementarity-determining regions (HCDR3), some over 70 amino acids in length. The unusual architectures of these antibodies enable them to easily access the CD4bs, supporting a model of steric blocking for conventional IgG molecules.

In contrast, highly variable regions occupy the more accessible regions of the trimer, providing yet another immunodominant diversion (Fig. 1C). Within a few months of infection patients readily develop autologous nAbs (i.e. antibodies capable of neutralising only the strain they were raised against), often directed at the variable loops 1, 2 and 3 (V1/V2 and V3) [59, 60]. Antibody-mediated selection pressure, combined with an error-prone viral reverse transcriptase, rapidly drives viral escape and results in extreme diversity [60]. Indeed, the genetic diversity of HIV-1 within an infected individual is comparable to the global genetic diversity of influenza in one year [46]. Thus, an effective vaccine against HIV-1 must not only induce nAbs, but antibodies with sufficient breadth of activity to neutralise the majority of circulating strains (bnAbs).

The development of bnAbs is dependent on the activation of the appropriate naïve B cells by engaging their B cell receptor (BCR), i.e. the precursor bnAb, prior to affinity maturation of the B cell in the germinal centre by somatic hypermutation. The resulting bnAbs are often significantly mutated from their germline-encoded BCR [33]. However, this process is hindered by the low affinity of so-called “germline”-bnAbs (gl-bnAbs) to the envelope spike. Consequently, many recombinant Env mimics also fail to bind gl-bnAbs [61-64], adding to the difficulties in eliciting bnAbs in a vaccine setting (Section 6.3).

1.2 The glycan shield
A contributing factor to the immunodominance of many of the non-neutralising and autologous neutralising epitopes, and the inability of Env to bind gl-bnAbs, is the presence of an array of N-glycans that mask much of the surface of the envelope spike (Fig. 1B). Each Env can have upwards of 90 potential N-glycosylation sites (PNGS), with glycans comprising approximately half the trimer’s mass...
The extensive N-glycosylation presents additional challenges for the host antibody response to overcome. The glycans are derived from the host’s own glycosylation machinery during Env synthesis and are therefore considered immunologically ‘self’. Auto-reactive B cells undergo strong negative selection during B cell development, constraining the development of potential anti-glycan antibodies [66]. Furthermore, glycoproteins tend to exist as heterogeneous populations, with a multitude of glycan structures decorating the same protein backbone, thereby potentially reducing the antigenicity of each individual glycoform [67]. Lastly, protein-glycan interactions tend to have low binding affinities and often require multivalent interactions to overcome this [68]. The heavily glycosylated outer domain of gp120 has been dubbed the ‘silent face’ of HIV-1, due to the previous lack of antibodies described against this region [69].

The role of glycans in protecting HIV-1 from neutralising antibody responses has been well documented. The glycan shield constantly evolves to escape the host immune system, with the addition and deletion of glycan sites frequently used by the virus to escape nAb responses [70-72]. The ‘evolving glycan shield’ escape response is typified by the N332 glycan, which has been reported to shift from the N334 position and back again after the appearance of N332-dependent nAbs [70]. Furthermore, transmitted/founder (TF) viruses typically have fewer PNGS than chronic isolates [72-75]. While this suggests there may be a fitness advantage to having fewer PNGS, a balance must then be struck between maintaining viral fitness and protecting vulnerable epitopes from the emerging nAb response. More recently a study by Wagh et al. reported that the addition of PNGS to fill holes in the glycan shield in vivo resulted in increased resistance to autologous nAbs [76]. Indeed, many in vitro studies have also reported the increased susceptibility of the virus to neutralisation upon removal of PNGS [77-81], which can often be rationalised by clashes observed between nAbs and glycans in structural studies [82, 83].

Despite the abovementioned challenges, approximately a third of infected individuals develop some level of bnAbs after a few years of infection [11-13]. Many bnAbs are able to either penetrate the glycan shield to bind protein surfaces or directly bind to Env glycans. Thus, while glycan shielding remains a potent immune evasion strategy, the discovery of numerous glycan-binding bnAbs, has highlighted the glycan shield as part of an attractive target for vaccine design [84]. The importance of the glycan shield in HIV-1 vaccine design has recently been underlined by Wagh et al., who observed that the development of bnAbs in infected individuals correlated with the completeness of the glycan shield at transmission [76].
Although bnAbs are increasingly being isolated from infected individuals, we are yet to elicit them in a vaccine setting in humans. Early vaccination strategies using recombinant, monomeric gp120 failed to confer protection [85-87], presumably due to the elicitation of antibodies directed against the aforementioned immunodominant non-neutralising epitopes (and/or the absence of gp41 and quaternary epitopes) [88]. The focus of HIV-1 vaccine research has now shifted to include the production of trimeric Env immunogens that display the majority of bnAb epitopes while minimising non-neutralising epitopes as much as possible, with the hope that these will be better able to induce a bnAb response [39]. However, the envelope spike is inherently unstable, reflecting its need to undergo substantial conformational changes during viral and host cell fusion. This has made the design of native-like immunogens particularly challenging (for an extensive review on the history and design of native-like Env trimers, see Sanders and Moore, 2017 [39]).

The desire to remove the transmembrane region of the protein in order to generate soluble mimics often amplified trimer instability. Initial efforts to stabilise the trimer involved the removal or inactivation of the furin cleavage site to prevent gp120-gp41 dissociation, and/or the introduction of a trimerisation domain at the C-terminus to prevent separation of the three gp120-gp41 heterodimers [89, 90]. Although these approaches usually generated trimers, there was often an abundance of monomers, dimers, and higher-molecular weight aggregates, owing to the inappropriate formation of intermolecular disulphide bonds [91, 92]. These trimer constructs often displayed aberrantly folded gp120s due to intramolecular disulphide bond scrambling and non-native like quaternary structures as judged by peptide mapping, negative-stain electron microscopy (EM), hydrogen-deuterium exchange mass spectrometry and glycosylation analysis (Section 4.4). Perhaps most importantly, these trimers display non-native-like antigenicity [91-95] and are commonly referred to as ‘pseudotrimers’ to reflect their various non-native properties [39]. With hindsight it was therefore unsurprising that immunisation studies with pseudotrimers, not unlike monomeric gp120, failed to elicit nAbs with sufficient breadth or potency [96-98].

2. Protein mimicry in vaccine design

The development of the SOSIP.664 platform transformed the field of native-like HIV-1 trimer design (Fig. 1C) [39]. These constructs retained the furin cleavage site, optimised for efficient cleavage, and relied on the introduction of a disulphide bond (‘SOS’) to covalently link the gp120 and gp41 subunits [99]. An additional point mutation in the gp41 subunits (I559P, ‘IP’) strengthened interactions between the three heterodimers by trapping the trimer in its pre-fusion conformation [100, 101], while truncation before the transmembrane region (‘.664’) ensured the solubility of the trimers and reduced aggregate formation [102]. A further point mutation (T332N) introduced a glycan site that
contributes to the epitope of many bnAbs. This format was first successfully applied to a clade A strain, BG505 [54], with the resulting ‘BG505 SOSIP.664’ trimers exhibiting both native-like structure [103] and antigenicity [54].

In recent years, an arsenal of native-like trimers has been produced. The ‘SOSIP’ format has since been applied to multiple strains, and mosaic and consensus sequences, though these often required further stabilising point mutations and disulphide bonds (either between gp120-gp41 and/or between protomers) [104-112]. Other native-like trimer formats focused on eliminating the requirement for furin cleavage (and therefore the need to co-express Env immunogens with furin encoding plasmids) in an attempt to simplify protein production strategies and DNA based vaccines.

Native flexibly linked (NFL) and single-chain (SC) trimers achieved this feat by replacing the furin cleavage site with a flexible Gly-Ser linker [113, 114]. While the NFL and SC constructs both relied on the I559P point mutation to maintain the pre-fusion conformation, UFO (uncleaved prefusion-optimised) constructs achieved this through a computational redesign of the HR1 (heptad repeat 1) region [115]. The immunogenicity of many of the above mentioned native-like trimers have now been investigated in animal models [104, 107, 116-121], and have been reviewed by Sanders and Moore, 2017 [39].

3. Display of bnAb epitopes by viral spike mimetics

Native-like trimers are often assessed by their ability to bind bnAbs and not non-neutralising antibodies. The development of native-like soluble trimers enabled the characterisation of many bnAb epitopes through various biophysical techniques, such as X-ray crystallography and EM [122]. Broadly neutralising antibodies are generally categorised by their recognition of five distinct and largely conserved epitopes on the envelope spike: the CD4 binding site (CD4bs), the membrane proximal external region (MPER), and the N-glycans located at the gp120-gp41 interface, the outer domain of gp120, and on the V1/V2 loops at the trimer apex. However, recent advances in high throughput B cell screening have led to a dramatic increase in the identification of new bnAbs, and have subsequently revealed a continuum of epitopes spanning the entire surface of the trimer [34], including the majority of N-glycans (Fig. 2). The recent characterisation of bnAb VRC-PG05, which recognises glycans at positions N262, N295 and N448, could conceivably be the last class of bnAb to be identified as its discovery filled one of the few remaining gaps on the trimer surface [123].
The extent to which individual bnAbs depend on glycans for binding and neutralisation varies. At one extreme, 2G12 recognises an epitope comprised exclusively of N-glycans [124-130]. On the other hand, bnAbs against the CD4bs, for example, have evolved to either avoid or accommodate glycans that would otherwise occlude the underlying protein epitope [131]. Most bnAb epitopes occupy a midway point, with binding dependent on both protein and glycan components.

The abundance of glycan-targeting bnAbs has earmarked N-glycans as important components of a future HIV-1 vaccine. Thus, there is a need for a detailed definition of the precise composition of the glycan shield on both viral Env and candidate immunogens. Additionally, an understanding of the principles controlling glycosylation will help guide the design of immunogens that are able to effectively mimic viral Env glycosylation, and, hopefully, elicit a broadly neutralising anti-glycan response.
3.1 Structure and development of broadly neutralising antibodies

The development of bnAbs requires repeated rounds of viral escape and antibody maturation and can therefore take several years of infection to arise [70, 132, 133]. Their slow development may reflect the fact that bnAbs often depend on unusual antibody features in order to overcome the aforementioned challenges associated with the development of glycan recognition and neutralisation breadth.

The glycan-targeting antibody, 2G12, overcomes low affinity protein-glycan interactions by exhibiting a unique domain-exchanged structure. Here, two heavy chain variable regions are exchanged to create a single Fab₂ [127]. The resulting structure has an additional antigen binding site at the interface of the two arms, which allows for the binding of four Env glycans with high avidity [126, 127].

Many bnAbs, particularly those targeting the V1/V2 and outer domain glycans, contain very long HCDR3 sequences [134-139]. While the average length of human HCDR3 is 13 residues long [140], bnAb HCDR3s can contain upwards of 30 residues. This is particularly true of bnAbs targeting the glycans of the outer domain and V1/V2 loops, where long HCDR3 sequences allow the antibody to penetrate the glycan shield and make contact with the underlying protein surface. The HCDR3 of PGT145, for example, contains 33 residues, and penetrates the N160 glycan triad at the apex of the trimer to contact underlying protein residues from all three protomers [134]. For outer domain-targeting bnAbs (e.g. PGT128 and PGT135), long HCDR3s allow penetration of the glycan shield to access the base of the V3 loop [136, 137]. In both examples, at least one glycan makes extensive interactions with the bnAb binding site. A similar phenomenon has been observed for bnAbs targeting the MPER, namely 4E10, where a long (20 residue) hydrophobic HCDR3 allows for contact with the lipid membrane as well as gp41 protein surface [139]. Often long HCDR3s are accompanied by post-translational modifications, such as tyrosine sulphation, which have also been reported to contribute to binding and neutralisation [135].

Another feature present in many bnAbs is extensive somatic hypermutation [33]. Classically, antibodies accumulate such mutations in the antigen-contacting complementarity-determining regions (CDRs). However, bnAbs often require somatic mutations to the more conserved framework regions [141]. Although there is evidence that higher levels of somatic hypermutation correlate with increased breadth and potency in some bnAb lineages [142], this may be a consequence of the length of time of development rather than a necessity of bnAb activity. For instance, partially germline reverted forms of VRC01 and 10E8 bnAbs are still broad and potent neutralisers [143].
Poly- and auto-reactivity are also frequently associated with bnAbs [144-146]. These observations are somewhat unusual as both characteristics are negatively selected for during B cell development [66]. It has been hypothesised that conserved HIV-1 epitopes may mimic host proteins in order to avoid host antibody responses through the down-regulation of relevant B cells by host tolerance mechanisms [144, 145, 147]. A similar host immune evasion strategy has been reported for the pathogen Campylobacter jejuni, whose lipooligosaccharide (LOS) mimics the gangliosides of the host's peripheral nervous system [148]. Despite these barriers, poly- and auto-reactive bnAbs are not uncommon, though their development may be dependent on prolonged exposure to the antigen, as several B cell tolerance checkpoints may need to be overcome. The elicitation of bnAbs through typical vaccination strategies may therefore prove challenging.

4. Glycan mimicry in vaccine design

The observation that glycan-dependent bnAbs can bind to soluble mimetics of the viral spike indicates that the glycans displayed by such immunogens are largely tolerated within the bnAb glycan specificity profile. However, given the nature of bnAbs, which have evolved to tolerate microheterogeneity of glycans, cross-reactivity does not necessarily indicate that the target glycans are precisely conserved between immunogen and virus. For example, some glycan specific bnAbs recognise the largely invariant base of the glycan [134, 149]. Furthermore, failure of an immunogen to bind a bnAb could be symptomatic of either a failure of protein mimicry, glycan mimicry, or both. For these reasons, there is significant interest in defining the glycosylation of both target viruses and candidate immunogens. Finally, detailed information about glycosylation will also help define the immunological ramifications of particular immunogen and viral glycoforms that may go beyond simply the display of particular glycan structures at a particular site. For example, are some glycoforms more inflammatory or immunogenic than others [150]?

4.1 Understanding Env glycosylation processing

The analysis of HIV-1 glycosylation is particularly challenging given the extensive heterogeneity displayed by glycoproteins, combined with the large number of PNGS on Env. Furthermore, isolating virally-derived Env in sufficient quantities for glycan analysis has proved difficult. Early glycan analyses were therefore generally performed on recombinant, monomeric gp120 [151-157], or trimeric Env constructs derived from pseudovirions [151, 158-160], membrane-associated trimers [161, 162], and recombinant, soluble trimers [45, 94, 121, 162-170]. These analyses soon revealed key aspects of Env glycosylation processing. Firstly, the Env glycan shield is heterogeneous. The gp120 subunit alone can contain upwards of 50 different glycan structures [45], attributed to the large number of glycan processing enzymes possessed by mammalian cells. Despite this heterogeneity, there always exists a
substantial population of under-processed, oligomannose-type glycans (Man$_{5-9}$GlcNAc$_2$). This glycan signature arises through steric constraints within the Env glycan shield that impede the actions of some of the host glycosylation enzymes (Section 4.3 and 4.4) [156, 158, 159, 171].

Many of the above glycan analyses focused on the analysis of enzymatically released N-glycans. While this provides a useful readout of the overall glycosylation profile, no site-specific information can be gleaned. Such information, which involves the analysis of protease-digested glycopeptides, is valuable in order to elucidate the precise composition of bnAb epitopes. The development of the SOSIP.664 platform, combined with an advancement of mass spectrometric and chromatographic technologies enabled more in-depth, site-specific glycan analyses.

We have previously published a quantitative, site-specific glycan analysis of the recombinant BG505 SOSIP.664 trimers [45]. This revealed that the oligomannose signature observed on Env was largely accounted for by several PNGS containing solely under-processed, oligomannose-type glycans, usually dominated by Man$_{9}$GlcNAc$_2$. The remaining sites contained either processed, complex-type glycans, or a mixed population. Mapping this information onto the structure of BG505 SOSIP.664 illustrated how large regions of oligomannose-type glycans span across the outer domain of gp120, while complex-type glycans occupied sites at the periphery of the trimer, particularly on the gp41 subunits (Fig. 3).
Figure 3. Principles controlling Env glycosylation. Site-specific glycan analysis of recombinant BG505 SOSIP.664 [45, 163, 168, 172] and virally-derived BG505 Env [173, 174] has revealed clusters of glycans displaying under-processed, oligomannose-type glycans (green). These are largely located on the outer domain of gp120 (forming the intrinsic mannose patch) and at the trimer apex and protomer interfaces (forming the trimer associated mannose patch). Model according to Figure 1. (A) The quaternary protein structure of native-like trimers imposes steric constraints on the host’s glycosylation enzymes, resulting in an increase in the amount of oligomannose-type glycans at sites near the protomer interfaces (red/orange), compared to that of monomeric or non-native trimeric Env [168]. (B) Irrespective of the mature Env peptide sequence, the presence of a signal peptide (SP) from a transmitted/founder (TF) viral isolate results in an increase in oligomannose-type glycosylation, while a chronic-stage signal peptide results in increased complex-type glycosylation. The signal peptide influences Env trafficking, folding and retention through the ER [175]. (C) Modelling the ER α-mannosidase I (cyan, PDB: 5KIJ) on to its substrate glycan (green) reveals extensive clashes with neighbouring glycans (red) sufficient to explain the formation of the intrinsic mannose patch [171]. (D) It is hypothesised that membrane-bound Env constructs display elevated glycosylation processing as they exhibit a different topology relative to the membrane-bound enzymes compared to soluble constructs, which are released into the lumen of the ER [173, 174]. Schematic of a membrane-bound glycan processing enzyme, based on the structure of a sialyltransferase (PDB: 6APL). (E) While the processing of many Env glycans is limited by protein-directed steric constraints, the fate of others is dependent on the glycosylation enzymes possessed by the host cell. The gp41 from BG505 SOSIP.664 expressed in Chinese hamster ovary (CHO) cells displays increased sialylation compared to the same protein expressed in human embryonic kidney (HEK) cells [94].
The Paulson laboratory employed a complementary site-specific workflow to analyse BG505 SOSIP.664 trimers. By sequential digestion of glycopeptides with Endoglycosidase H (to cleave oligomannose-/hybrid-type glycans, leaving a GlcNAc residue) and Peptide-N-Glycosidase F in the presence of $^{18}O$-water (to cleave the remaining complex-type glycans and convert the Asn to an $^{18}O$-labelled Asp), they were able to generate novel mass signatures for each category of glycan: +203 Da for oligomannose-/hybrid-type, +3 Da for complex-type glycans, and +0 for unoccupied peptides. Their data were in very close agreement to our earlier analysis. This classification method is very powerful for low abundance samples (as heterogeneous glycopeptides become grouped as a single peptide), although information on the exact composition of glycopeptides is not captured.

In some instances, the detail obtained from intact glycopeptide analysis may explain < 100% neutralisation plateaus observed by some bnAbs [176, 177]. For instance, PGT135 is only able to neutralise around 85% of BaL pseudoviruses [176]. This bnAb recognises predominantly glycans at N332, N386, and N392 (Fig. 2) [137]. Site-specific analysis of the N392 PNGS on recombinant gp120$_{\text{BaL}}$ (subscript denotes strain) revealed the majority of glycans at this site to be Man$_9$GlcNAc$_2$ structures, with a secondary population of Man$_6$GlcNAc$_2$ [176]. In line with this, a crystal structure of the PGT135 Fab bound to a gp120$_{\text{BaL}},$FL core shows the N392 site to be occupied by a Man$_6$GlcNAc$_2$ structure [137]. Modelling an additional mannose residue to this glycan (to give Man$_9$GlcNAc$_2$) revealed steric clashes with the bnAb [137]. Furthermore, it has been shown that PGT135 is unable to neutralise pseudovirus displaying predominantly Man$_9$GlcNAc$_2$ structures [137]. Thus, Man$_6$GlcNAc$_2$ at the N392 site seems unoptimal for PGT135 binding, and the site-specific presence of this structure may account for the observed neutralisation plateaus.

4.2 Immunogen mimicry of viral glycosylation

Recent progress in the production and purification of HIV-1 virions has enabled the glycosylation analysis of virally-derived Env. The Dell laboratory presented a qualitative analysis of gp120$_{\text{BaL}}$ glycosylation derived from virions produced in a human lymphoid cell line [178]. In line with analysis of SOSIP.664 trimers, 15 of the 24 PNGS contained solely oligomannose-type structures, although the relative abundances of each glycoform were not determined. The remaining nine contained either complex-type structures or mixed populations [178].

We have since performed site-specific glycan analysis on gp120$_{\text{BG505}}$ isolated from virions expressed in a similar lymphoid cell line [173]. This allowed for a comparison with BG505 SOSIP.664 trimers expressed in both human embryonic kidney (HEK) cells (as per previous analyses), and under equivalent conditions to current Good Manufacturing Practices (cGMP) in a Chinese hamster ovary (CHO) cell line. The analysis revealed that sites occupied by exclusively oligomannose-type glycans on
virally-derived Env were largely conserved on recombinant SOSIP.664s. This is reassuring considering a large proportion of glycan-targeting bnAbs recognise the oligomannose-type glycans at these sites (Fig. 2). However, there were some key differences between virally-derived and recombinant Envs. Namely, virally-derived Envs displayed increased levels of glycosylation processing, both in terms of the relative amount of complex-type glycosylation, and also the structures present (i.e. complex-type glycans on virally-derived Envs were more branched). This discrepancy was generally attributed to several complex-type PNGS on virally-derived Env that contained mixed populations of glycans on the SOSIP.664 trimers. We propose this is due to the membrane-bound nature of virally-derived Envs and will discuss this further in Section 4.6.

The Paulson laboratory have also compared the glycosylation of virally-derived Env from three strains (JR-FL, BG505, and B41), produced in peripheral blood mononuclear cells (PBMCs), with their corresponding SOSIP.664 trimers, produced in HEK cells, using their aforementioned site-specific classification method [174]. Their results were in general agreement with our observations that mixed sites on SOSIP.664 trimers tended to be fully processed on the virally-derived Env. Thus, while there is some differential processing between viral and recombinant Env, many key bnAb epitopes are conserved. In terms of immunogen design, it is currently unknown whether absolute mimicry of viral glycosylation is required. This is likely to depend upon the specific epitope targeted, for example if the epitope mainly comprises the conserved base of the glycan.

4.3 Glycan clustering and the intrinsic mannose patch

This abundance of oligomannose-type glycans on Env is somewhat unusual given the virus derives its glycan shield from the host cell glycosylation machinery. This typically follows a highly ordered pathway whereby oligomannose-type precursors are trimmed and rebuilt as complex- and/or hybrid-type glycans (Fig. 4). Two hypotheses existed as to why Env glycosylation diverges from that typically observed on mammalian glycoproteins. Either a proportion of Env glycoproteins are exiting the glycosylation pathway before encountering the later enzymes, or steric constraints exist within Env that are preventing complete processing by the earlier enzymes, or a combination of the two. Several pieces of evidence support the model of steric hindrance.
Figure 4. Overview of the mammalian N-glycosylation pathway. The envelope spike is extensively glycosylated by the host cell, which typically follows a highly ordered pathway. As the protein is translated, a Glc3Man9GlcNAc2 (Glc = glucose, Man = mannose, GlcNAc = N-acetylglucosamine) precursor is transferred en bloc to Asn residues within the N-glycan consensus sequence Asn-X-Thr/Ser (where X is any amino acid except Pro). As the protein is folded the three terminal glucose residues are removed to give rise to a glycoprotein displaying homogenous Man5GlcNAc2 structures. This is then further trimmed by endoplasmic reticulum (ER)- and Golgi apparatus-resident α-mannosidases to give rise to Man3GlcNAc2. Steric constraints within Env limit the actions of these early enzymes resulting in a population of under-processed oligomannose-type glycans. The addition of a β1-2-linked GlcNAc residue to Man3GlcNAc2 structures initiates cell-specific diversification to a variety of hybrid- and complex-type structures, through additional processing and/or trimming. α-mannosidase I and II, GnT I = N-acetylglucosaminyltransferase I, Gal = galactose, Fuc = fucose, Neu5Ac = N-acetylneuraminic acid (sialic acid). Glycan structures are depicted in symbols according to the Consortium for Functional Glycomics nomenclature, with linkage information according to Oxford nomenclature, as shown in the key.
Firstly, a proportion of oligomannose-type glycans are resistant to endoplasmic reticulum (ER) α-mannosidase I (the enzyme responsible for trimming Man$_9$GlcNAc$_2$ to Man$_8$GlcNAc$_2$, Fig. 4) digestion in vitro [159]. Expressing gp120 with kifunensine (an α-mannosidase I inhibitor) generates a glycoprotein bearing almost exclusively Man$_9$GlcNAc$_2$ structures, replicating the immature glycoprotein found in the early ER. Kinetic analysis of the hydrolysis of Man$_9$GlcNAc$_2$ to Man$_8$GlcNAc$_2$ revealed approximately half of the Man$_9$GlcNAc$_2$ to be rapidly trimmed to Man$_8$GlcNAc$_2$, with the remaining Man$_9$GlcNAc$_2$ processed at a much slower rate. However, there remained a proportion of Man$_9$GlcNAc$_2$ (~30%) that could not be hydrolysed, even after exhaustive digestion. The resistance of these glycans to digestion in vitro supports the model of steric hindrance, likely caused by the high density of glycans on gp120 preventing α-mannosidases accessing their substrate glycans.

Each gp120 subunit contains between 18 and 33 (median 25) PNGS [46]. To elucidate whether glycan density is the limiting factor preventing complete processing of the glycan shield, the number of PNGS on the outer domain of gp120 were correlated with the abundance of oligomannose-type glycans on sequences isolated from an individual over the course of infection, and on a cross-clade panel of 29 strains [157]. In both instances, a strong correlation was observed. This correlation was not seen when comparing the abundance of oligomannose-type glycans with the total number of PNGS on gp120 for the cross-clade panel (though it was for the infected patient sequences) suggesting that the incomplete processing observed on gp120 is driven by local glycan density, rather than overall glycan number. Similarly, Stewart-Jones et al. observed the number of glycan processing steps to be significantly lower for ‘crowded’ glycans (those with more than 15 PNGS within a 50Å radius) than for ‘dispersed’ glycans (those with fewer than 15 PNGS within a 50Å radius) [149].

The extent to which individual glycans contribute to the steric hindrance of glycosylation enzymes was investigated by the systematic site-directed mutagenesis of all the PNGS on gp120$_{bas}$, by mutating each Asn within the glycosylation sequon to Ala [156]. Although the removal of individual sites did not have a severe effect on the abundance of oligomannose-type glycans, several site deletions resulted in a larger than expected loss of Man$_9$GlcNAc$_2$, often accompanied by a compensatory increase in the lower oligomannose species, Man$_5$-8GlcNAc$_2$. The glycan sites accounting for the largest decrease in Man$_9$GlcNAc$_2$ generally mapped to the outer domain of gp120, further supporting a model whereby localised glycan clustering is sterically hindering early glycosylation processing. Thus, the loss of a PNGS within such clusters would increase the accessibility of neighbouring glycans so that multiple nearby glycans exhibit increased processing. In further support of this hypothesis, it has recently been confirmed by site-specific glycan analysis that the disruption in glycan processing upon the loss of a PNGS is largely limited to those sites adjacent to the missing glycans [179].
Thus, while the glycan structures present on most glycoproteins are determined by the host cell, HIV-1 is able to partly modulate its glycosylation processing through the number and position of PNGS, as determined by the viral sequence. The number of PNGS on Env are conserved both across clades and longitudinally throughout infection [157, 158]. The resulting high glycan density on Env (and gp120 in particular) limits the actions of α-mannosidases and results in a large population of oligomannose-type glycans. This principle is illustrated in a model of the ER α-mannosidase I enzyme binding a substrate Man$_9$GlcNAc$_2$ glycan on the outer domain of gp120, which reveals extensive clashes with the surrounding glycans [171]. A similar phenomenon has been observed on the glycoprotein complex of Lassa virus, which also exhibits high localised glycan density resulting in under processed oligomannose-type clusters [180]. The oligomannose-type glycans of gp120 are therefore an inherent feature of the glycoprotein, known as the ‘intrinsic mannose patch’ (IMP), and to some extent all Env constructs will display this glycan signature (Fig. 5). Encouragingly, many of the aforementioned glycan-targeting bnAbs recognise the oligomannose-type glycans of the IMP. The overall resilience of the oligomannose population to sequence variation, and its presence on both immunogens and viral Env, supports this feature as a conserved target for vaccine design.
Figure 5. Glycan clustering and protein structure limit glycosylation processing on HIV-1.

Quantitative glycan analysis of monomeric gp120, pseudotrimers, native-like trimers and virally-derived Env. For comparison, data for both gp120 and gp140 (gp120 + truncated gp41) from native-like trimers have been included, only data for virally-derived gp120 was available [168, 173]. Each construct is based on the BG505 sequence containing the T332N mutation. Glycans were enzymatically released, fluorescently labelled, and analysed by hydrophilic interaction liquid chromatography-ultra performance liquid chromatography. Oligomannose-type glycans were quantified by their susceptibility to digestion with Endoglycosidase H. The chromatograms reveal a population of oligomannose-type glycans (green) intrinsic to all Env constructs, termed the intrinsic mannose patch (IMP). Only native-like trimers and virally-derived Envs display the additional trimer-associated mannose patch (TAMP) signature, attributed to additional steric protection from processing.

4.4 Protein structure and the trimer associated mannose patch

As the focus of immunogen design shifted from monomeric gp120 toward trimeric Env, so did the requirement to categorise glycans from native-like immunogens. It had been observed that trimeric Envs often displayed even higher levels of under-processed oligomannose-type glycosylation [94, 158, 159, 168]. The discrepancy in the abundance of oligomannose-type structures between monomeric and trimeric Env suggests that these structures are derived from an additional mechanism to the one driving the formation of the IMP. It was hypothesised that the ‘trimer associated mannose patch’ (TAMP) arose due to glycan-glycan and glycan-protein interactions at the protomer interface further sterically restricting processing enzymes [41, 158, 159].
Quantitative glycan analysis soon revealed that the TAMP glycan signature was only present on native-like trimers, with non-native pseudotrimers often displaying glycosylation patterns similar to that of monomeric gp120 (Fig. 5) [92, 94, 168]. Negative-stain EM revealed that non-native pseudotrimers often form open, irregular structures, in stark contrast to the well-ordered, propeller shaped format of native-like trimers [92, 94, 171]. This structural difference is sufficient to explain the discrepancy in glycosylation, as open, irregular structures lack the additional steric protection of native-like trimers and allow for more complete glycan processing. Thus, although most Env proteins have a population of oligomannose-type glycans atypical of mammalian glycosylation, only native-like trimers display the additional TAMP signature. As such, glycosylation profiling is becoming a widely-adopted tool that can be used to readily distinguish between native-like and misfolded trimers and guide the selection of immunogens [181].

Site-specific glycan analysis was applied to sequence-matched gp120<sub>BG505</sub>, uncleaved pseudotrimers, and native-like, BG505 SOSIP.664 trimers. The results revealed for the first time the individual PNGS that contribute to the TAMP. As hypothesised, sites located at the protomer interfaces on the native-like BG505 SOSIP.664 trimers displayed restricted glycan processing compared to both the pseudotrimers and monomeric gp120 (Fig. 3A). This was most significant at sites N156, N160, N197 and N276 which had > 40 percentage point increase in oligomannose-type glycans on BG505 SOSIP.664 compared to gp120 [168].

Crucially, in support of the native-like configuration of SOSIP.664 trimers, the TAMP signature is present on both virally-derived Envs and native-like trimers (Fig. 5), though it is somewhat smaller on viral Env [173]. This is likely due to ‘TAMP sites’ N197 and N276 displaying predominantly complex-type glycosylation on viral Env [173, 174]. The N156 and N160 ‘TAMP sites’ are, however, conserved oligomannose-type sites in most instances [173, 174]. The N160 site has been reported as a complex-type site in JR-FL virions, highlighting potential strain-specific differences [174].

4.5 Cell-specific glycosylation

While the abundant population of oligomannose-type glycans observed on Env are attributed to steric constraints imposed by the protein itself, the more accessible regions of the glycoprotein are subject to processing by the host cell’s glycosylation pathway. Such processing is cell type-specific, dependent on the repertoire of glycosidases and glycosyltransferases expressed by the host cell [182].

A comparative analysis of BG505 SOSIP.664 derived from CHO cells and HEK cells illustrates the nature of protein- and cell-directed Envelope glycosylation [94]. Both CHO and HEK-derived material display similar levels of protein-directed glycosylation: oligomannose-type glycans account for 55 and 56% of the total glycan pool, respectively, with individual oligomannose species exhibiting
similar distributions [94]. In contrast, the remaining complex-type glycans present considerable cell-
specific differences. This is particularly evident on the gp41 subunit, with CHO-derived material
containing substantially more sialylated structures than HEK-derived BG505 SOSIP.664 (Fig. 3E) [94].
The abundance of sialic acid is known to impact the antigenicity and immunogenicity of Env. Kong et
al. observed that gp120 expressed in an insect cell line modified to impart asialylated mammalian-
type glycans was significantly more immunogenic than sialylated gp120 expressed in HEK cells [150].

A comparison of virally-derived gp120 produced in PBMCs and gp120 from pseudovirus
produced in HEK cells revealed similar cell-specific differences [183]. Namely, sialylated glycans from
material produced in PBMCs displayed primarily α2-6-linkages, whereas HEK-derived material
displayed only α2-3-linked structures [183]. It can be noted that α2-6-linked sialylated glycans have
previously been associated with anti-inflammatory effects, thus there may be potential immunological
consequences of these structures [184].

Like oligomannose-type glycans, complex-type glycans have emerged as important in regard
to the formation of bnAb epitopes. Some bnAbs, such as PGT121 display promiscuous recognition of
both oligomannose- and complex-type glycans in glycan arrays [185]. Antibodies such as PGT151 are
dependent on tri- and tetra-antennary structures at the gp120-gp41 interface [186], while α2-6-linked
sialylated hybrid- and complex-type glycans have been implicated in the development and binding of
apex-targeting bnAbs such as PG16 and the CAP256-VRC26 lineage [187-189]. Interactions between
the sialic acid residues of Env and Siglecs (sialic acid-binding immunoglobulin-like lectins) have also
been observed to play an important role in the infection of macrophages, which express low levels of
cell-surface CD4 [190]. Thus, cell-directed glycosylation can play a role in both viral infectivity and
bnAb binding, and understanding the factors influencing complex-type glycosylation may have
implications in guiding immunogen design, for instance when choosing expression cell lines.

4.6 Membrane tethering

The majority of glycan analyses have been performed on soluble Env constructs due to the difficulties
associated with expressing and purifying full-length, membrane-bound Env. Until recently, the impact
of membrane tethering on Env glycosylation had remained largely unaddressed. Analyses of virally-
derived Envs, while performed on membrane-bound material, generally did not control for expression
cell-line, known to influence glycosylation processing [173, 174]. However, recent studies by
Rantalainen et al. and Cao et al. directly compare the glycosylation of membrane-bound Envs with
their corresponding soluble SOSIP.664 trimers expressed in the same cell lines. The differences
observed between membrane-bound Env and SOSIP.664 are reminiscent of the differences observed
between viral Env and SOSIP.664, in that sites containing mixed populations of glycans on the soluble
SOSIP.664 trimer typically displayed only fully processed, complex-type glycan structures on the full-length construct [191]. It was hypothesised that the more complete processing observed on full-length constructs is due to membrane-bound Env being kept in closer proximity to the membrane-bound glycosylation enzymes throughout the ER and Golgi (Fig. 3D), as opposed to soluble Env which is released into the lumen [173, 174]. In line with this hypothesis, membrane-bound CD59 also displays increased processing compared to its soluble counterpart [192].

Alternatively, it has been hypothesised that the closer proximity of the membrane to full-length constructs, particularly to gp41, would pose additional steric constraints on Env glycosylation processing [171]. In support of this hypothesis, Panico et al. observed exclusively oligomannose-type glycans at the membrane-proximal N88 site on virally-derived Env [178]. A caveat of all the above studies is that full-length constructs do not contain the SOSIP.664 stabilising mutations, which may influence glycosylation via a separate mechanism, for example, the speed at which Env transits through the ER.

4.7 Signal peptide

Like many membrane proteins, immature Env contains a signal peptide at the N terminus responsible for directing the nascent peptide to the ER, which is subsequently cleaved off prior to transport of the maturing peptide through the ER and Golgi (Fig. 3B). The signal peptide strongly influences the processing of Env as it transits through the ER, impacting factors such as trafficking to the ER, the rate of signal peptide cleavage, and the retention time of Env within the ER [193-198]. Natural variation exists within the signal peptides of Env. For example, TF viral isolates often over represent His at position 12 of the signal peptide, a signature not usually observed in chronic phase viral isolates [198].

Given the influence of the signal peptide on the molecular biology of Env, the laboratory of Fauci sought to determine the impact of the natural variation observed between the signal peptides isolated from TF viruses and from a chronically infected patients on the glycosylation, structure and antigenicity of Env. By creating four constructs containing either the gp120 sequence isolated from a TF or chronic virus, in combination with either their natural signal peptide, or that of the other virus (Fig. 3B), Yolitz et al. assessed glycosylation and antigenicity by binding to various lectins and bnAbs [175]. They found that despite gp120 sequences encoding for the same mature protein, the presence of a signal peptide from the TF viral isolate resulted in an increase in oligomannose-type glycosylation (as judged by binding to the *Narcissus pseudonarcissus* lectin and bnAb 2G12, specific to α-linked mannose residues), while the presence of a signal peptide from chronic-stage virus increased the amount of complex-type glycosylation (as judged by binding to *Ricinus communis* lectin which preferentially binds structures terminating in galactose).
The results of Yolitz et al. are in line with the glycan analyses of Go et al., who observed higher levels of oligomannose-type glycosylation on TF viruses than those isolated from chronically infected patients [165]. However, a similar comparison of SOSIP.664 trimers isolated from early and late time points by the Paulson laboratory showed very similar glycan processing [191]. Nevertheless, the findings of Yolitz et al., implicate the signal peptide, a domain that does not appear in the mature protein, as a regulator of glycosylation processing.

4.8 Occupancy
Potential N-glycosylation site occupancy has recently emerged as an important aspect of Env glycosylation. The Paulson laboratory employed site-specific glycan classification analysis to BG505 SOSIP.664 trimers to assess PNGS occupancy [163]. They found that overall occupancy was very high: all but four of the 28 PNGS (V1/V2 and gp41 sites N190, N197, N618, and N625) were > 90% occupied and none were < 50% occupied [163]. Extension of their method to native-like trimers derived from different strains (JR-FL, B41, CRF02_AG_250, 327c, PC64) confirmed that the V1/V2 and gp41 sites were most susceptible to under-occupancy [174, 191]. The reports concerning occupancy of gp41 sites, particularly N625, seem to vary, perhaps due to methodological or sample variation. For example, Guttman et al. have reported occupancy < 20% at this site in BG505 and KNH1144 SOSIP.664 trimers [199] which is also compatible with the known epitope of the 35O22 bnAb [82].

Interestingly, virally-derived and membrane-bound Env trimers are generally more occupied than their SOSIP.664 counterparts [173, 174, 191]. Although the mechanism through which PNGS under-occupancy arises is not yet understood, we hypothesise that the codon-optimisation process, used to increase the yield of recombinant SOSIP.664 trimers, may increase the rate of protein translation and folding and reduce the chance of an N-glycan being attached [173]. This may be particularly true of glycans in the V1/V2 region where several PNGS are in close proximity.

4.9 Considerations for immunogen manufacture
All of the above factors influence the glycosylation processing of Env to varying degrees. Although the extent to which Env immunogens must mimic the native viral envelope spike is not yet known, there are nevertheless important considerations for immunogen design and manufacture. In particular, expression cell-type and PNGS under-occupancy may yet emerge as crucial aspects in the cGMP production of Env immunogens, as under-occupancy may introduce unfavourable, immunodominant autologous nAb epitopes. This is particularly relevant in light of Wagh et al., observing increased neutralisation breadth in individuals infected with TF viruses with more intact glycan shields [76].
5. Understanding the nature of glycan-dependent bnAb epitopes

The structural characterisation of multiple bnAbs in complex with Env by both X-ray crystallography and EM has enabled a detailed description of the epitopes of many bnAb (Table 1) [200, 201]. Furthermore, it has revealed many of the unusual antibody features required for broad neutralisation. Nevertheless, there remained many unknowns about the precise nature of glycan epitopes. This is, in part, due to difficulties in resolving extremely heterogeneous glycans with structural techniques that rely on the averaging of many molecules to reveal the consensus structure [122]. Glycan arrays are also a useful tool in assessing the glycan-binding properties of bnAbs, though these too have limitations (e.g. if the epitope comprises protein components). Site-specific glycan analysis has gone some way to bridging these knowledge gaps. While there is generally good agreement of the precise structures occupying glycan epitopes between glycan analyses and structural studies/arrays, a few discrepancies remain. For example, there are several reports on the preference of apex-targeting bnAbs PG9 and PG16 for complex- or hybrid-type glycans (specifically α2-6-linked sialylated structures) at the N156 site [187, 188, 202]. However, these structures are seldom seen at this site in glycan analyses of recombinant, soluble trimers or virally-derived Env [45, 163, 173, 174].

Table 1. Examples of glycan-targeting broadly neutralising antibodies and their glycan epitopes.

<table>
<thead>
<tr>
<th>Epitope</th>
<th>bnAb (and key to Fig. 2)</th>
<th>PNGS</th>
<th>Glycan type</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGT130</td>
<td>N301, N332/N334</td>
<td>Oligomannose</td>
<td>[203, 204]</td>
<td></td>
</tr>
<tr>
<td>2G12</td>
<td>N295, N332, N339, N392</td>
<td>Oligomannose, specifically α1-2-linked motifs on Manα2-6GlcNAc2 structures</td>
<td>[124-130]</td>
<td></td>
</tr>
<tr>
<td>VRC-PG05</td>
<td>N262, N295, N448</td>
<td>Oligomannose</td>
<td>[123]</td>
<td></td>
</tr>
<tr>
<td>PGT135</td>
<td>N295, N332, N386, N392</td>
<td>Oligomannose</td>
<td>[137, 176, 204]</td>
<td></td>
</tr>
<tr>
<td>PGT128</td>
<td>N156, N295, N301, N332/N334</td>
<td>Oligomannose</td>
<td>[44, 136, 203, 204]</td>
<td></td>
</tr>
<tr>
<td>PGT122</td>
<td>N137, N156, N301, N332</td>
<td>Oligomannose</td>
<td>[149, 200, 204]</td>
<td></td>
</tr>
<tr>
<td>V1/V2 glycans</td>
<td>gp120/gp41 interface</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGT121</td>
<td>N137, N156, N301, N332/N334 Oligomannose or complex [142, 185, 203, 204]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BG18</td>
<td>N156, N332, N386, N392 Oligomannose, possibly complex at N156 [205]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-1074</td>
<td>N156, N301, N332 Oligomannose, possibly complex at N156 [185, 206]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGT145</td>
<td>N160 Oligomannose [134, 204]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH04</td>
<td>N160 Kifunensine abrogates neutralisation (can’t tolerate Man9GlcNAc2) [207]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGDM1400</td>
<td>N160 Kifunensine reduces/abrogates neutralisation (can’t tolerate Man9GlcNAc2) [208]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VRC38.01</td>
<td>N133, N156, N160 N-acetylglucosamine core of N133 [209]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PG9 and PG16</td>
<td>N173 (N156), N160 Oligomannose at N160, hybrid or bi-antennary structures at N156/N173, specifically containing α2-6-linked terminal sialic acids [187, 188, 202, 210-212]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAP256-VRC26 (lineage)</td>
<td>N156, N160 Oligomannose at N160, hybrid at N156, or bi-, tri-, tetra-antennary structures containing α2-6-linked terminal sialic acids [133, 189]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCT64 (lineage)</td>
<td>N156, N160 Oligomannose, preferentially Man9GlcNAc2 [213]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3BNC117</td>
<td>N197, N276 N-acetylglucosamine cores [134, 149]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGV04 (VRC-PG04)</td>
<td>N276, N301, N363, N386 Enzymatic de-glycosylation doesn’t affect binding (no strong glycan-dependence) [201, 214]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HJ16</td>
<td>N276 N-acetylglucosamine core [215-217]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IOMA</td>
<td>N197, N276, N363 Complex, minor contact with oligomannose glycan at N363 [206]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>179NC75</td>
<td>N276 Oligomannose or hybrid† [218]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VRC01</td>
<td>N276 N-acetylglucosamine core [131, 149]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8ANC195</td>
<td>N234, N276, N637 Kifunensine did not affect neutralisation (can tolerate Man9GlcNAc2) [219, 220]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Footnotes:**

† In Freund et al., 2015, the loss of binding of 179NC75 following expression of BG505 SOSIP.664 trimers with kifunensine was used to argue a dependency on complex-type glycans. However, they also report a loss of binding to gp120 after digestion with Endoglycosidase H (which cleaves oligomannose- and hybrid-type glycans). This suggests that 179NC75 recognises Man$_8$GlcNAc$_2$ or hybrid-type structures. In line with this, site-specific glycan analysis of BG505 SOSIP.664 trimers reports predominantly Man$_8$GlcNAc$_2$ glycans at the N276 site, with a small population of hybrid-type structures [45].

‡ Treatment with swainsonine (α-mannosidase II inhibitor) results in oligomannose- and hybrid-type glycosylation. Expression in GnTI$^{-/-}$ cells (deficient in N-acetylglucosaminyltransferase I) results in Man$_5$GlcNAc$_2$ structures only.

6. **Beyond antigenic mimicry**

Native-like trimers are excellent structural mimics of viral Env, display similar glycosylation profiles, and are capable of binding bnAbs. Yet immunogenicity studies with native-like trimers have only routinely elicited autologous Tier-2 nAbs and heterologous Tier-1 nAbs [107, 116-121]. Weakly neutralising heterologous Tier-2 antibodies have also been reported in rabbits [104]. With the exception of cows (whose antibodies naturally contain very long HCDR3), the potent bnAbs required for a protective vaccine have not yet been generated via immunisation with native-like trimers [58]. The elicitation of bnAbs against the envelope spike will inevitably be challenging. Broadly neutralising epitopes are broad due to their conserved nature, though conservation generally correlates with poor immunogenicity as otherwise the epitope would have been selected against by immune pressure. In this section, we will address some of the strategies aimed at increasing the immunogenicity of this target.

6.1 **Mimicry of glycan epitopes: chemical approaches**

The first of many glycan-binding bnAbs to be discovered was 2G12. To date, this the only bnAb identified to exclusively bind glycans. The 2G12 antibody recognises clusters of α1-2-linked mannose motifs present on the Man$_8$GlcNAc$_2$ structures of the IMP (Fig. 2) [124-130]. Glycan based vaccine
strategies have thus far centred on the design of immunogens that mimic the 2G12 epitope. Several groups have created immunogens displaying multivalent, chemically synthesised oligomannose-type glycans attached to various scaffolds, including: carbohydrate, cholic acid, cyclic peptide, dendrimer, DNA, and gold (reviewed in Wang, 2013 [226]). Although many of these constructs are able to bind 2G12, and could elicit antibodies with specificity to oligomannose-type glycans, none have induced antibodies capable of neutralising HIV-1. This may be due to the inability of these antibodies to recognise clusters of oligomannose-type glycans as they are presented on Env, perhaps due to their non-domain-exchanged structures. Domain-exchange (at least for 2G12) is a requirement for HIV-1 neutralisation [227].

The mimicry of other glycan-targeting bnAb epitopes has been explored with the design of peptide-based immunogens. Glycopeptides based on the V1/V2 region, containing the N156 and N160 PNGS, are capable of binding both apex targeting bnAbs and their unmutated common ancestors [202, 228, 229]. Similarly, glycopeptides based on the V3 region, containing the N332 PNGS, bound outer domain targeting bnAbs such as PGT128 [230]. In immunogenicity studies, V3 glycopeptides elicited antibodies capable of binding gp120 but unable to neutralise virus [230, 231]. The failure of antigenic mimics of bnAb epitopes to induce bnAbs further highlights the complicated relationship between the antigenicity and immunogenicity of these epitopes.

6.2 Mimicry of glycan epitopes: biological approaches

A potential failure of the above strategies may be that, while the oligomannose-type glycans of the HIV-1 glycan are not typically observed on mammalian glycoproteins, they are still fundamentally ‘self’ structures. As noted above, this poses a challenge for the development of glycan-targeting bnAbs, as cross-reactive B cells are likely to be deemed autoreactive and will be negatively selected for. The fact that bnAbs generally do not arise until after several years of infection could support the notion that they are operating at the edge of immunological tolerance, or may simply reflect the long maturation process. However, tolerance to self-structures can be broken. For example, in some cases, infection with C. jejuni elicits an antibody response against its bacterial lipooligosaccharides that is able to bind nearly identical structures on the gangliosides of the peripheral nervous system (Fig. 6A) [148]. The resulting autoimmune disease, Guillain-Barré syndrome, provides proof of concept that immunological tolerance to ‘self’ glycans can be broken by exposure to micro-organisms bearing similar structures [42].
Figure 6. Antigenic mimicry in autoimmune disease and HIV-1 vaccine design. (A) Antigenic mimicry of lipooligosaccharides (LOS) from Campylobacter jejuni causes Guillain-Barré syndrome, an autoimmune response against the GM1 ganglioside in the peripheral nervous system (PNS). (B) LOS from Rhizobium radiobacter Rv3 and Saccharomyces cerevisiae deficient in the Mnn1 gene display glycan structures terminating in α1-2-linked mannose residues, mimicking the 2G12 epitope on HIV-1. GalNAc = N-acetylgalactosamine, Glc = glucose, GlcN = glucosamine, KDO = 2-keto-3-deoxy-o-manno-octulosonic acid. Figure adapted from Scanlan et al., 2007 [42].

It was hypothesised that 2G12 may have originally evolved to recognise high-mannose structures on pathogens other than HIV-1 [42]. Scanlan explored this further by assessing the reactivity of 2G12 to various yeast species [146]. Yeast typically display high-mannose structures that comprise an α1-6-linked mannose backbone from which branch repetitive α1-2-linked mannose motifs (i.e. a key feature of the 2G12 epitope), sometimes capped with α1-3-linked mannose residues. Of the yeast species tested, 2G12 was able to bind to two, binding Candida albicans with similar affinities to that of gp120 [146, 232]. This interaction was glycan dependent as binding was inhibited by D-fructose, a known ligand for the 2G12 binding site. Additionally, a non-glycan-dependent control bnAb was
unable to bind any of the yeasts. Thus, the 2G12 antibody may equally well be described as an anti-
Candida antibody with cross-reactivity to HIV-1 [146].

Scanlan went on to formally elucidate the relationship between the antigenicity and
immunogenicity of yeast glycosylation [233]. Saccharomyces cerevisiae deficient in mnn1, the gene
responsible for the α1-3-linked mannose caps, displays mannans terminating in the α1-2-linked
mannose motif (Fig. 6B). Immunisation of rabbits with S. cerevisiae Δmnn1 elicited antibodies with
similar glycan specificities to 2G12. Furthermore, the sera displayed weakly neutralising activity
against HIV-1 [233]. The Doms laboratory also investigated using yeast as an HIV-1 immunogen. By
knocking-out three genes they produced S. cerevisiae displaying predominantly Man8GlcNAc2 glycan
structures. Although these yeasts now displayed a ‘self’ glycan structure, immunisation of rabbits
elicited oligomannose-specific antibodies, capable of binding monomeric gp120, but unable to
neutralise the virus [234]. As mentioned previously, an explanation for this apparent contradictory
result may be that while the elicited antibodies were capable of recognising isolated oligomannose-
type glycans, they were not able to recognise densely packed clusters of oligomannose-type glycans
as they are presented on Env. Thus, a logical next step may be to present immunogenic yeast glycan
structures in the context of Env, for instance, by expressing Env in yeast. The resulting immunogens
should display clusters of ‘non-self’ glycans capable of breaking immunological tolerance to the
clusters of oligomannose-type glycans present on Env.

A similar phenomenon has been observed by the Pantophlet laboratory. As per S. cerevisiae
Δmnn1 mannans, the lipooligosaccharides of Rhizobium radiobacter Rv3 bacteria contain a glycan
motif analogous to the 2G12 epitope on HIV-1 (Fig. 6B) [235, 236]. Immunisation of mice with R.
radiobacter Rv3 also generated antibodies capable of weakly cross-reacting with HIV-1 [235, 237],
representing another candidate micro-organism to investigate for a potential HIV-1 glycan-based
vaccine.

6.3 Mimicking bnAb development in natural infection

The development of bnAbs in a subset of infected individuals typically only occurs after a few years of
infection. Neutralising antibodies, however, arise early in infection, and exert a considerable selection
pressure on the virus. Repeated rounds of viral escape and antibody evolution ultimately drive the
development of nAb breadth [70, 132, 133]. The discovery that native-like trimers can induce
autologous Tier-2 nAbs could therefore be an important first step towards development of bnAbs.
Accordingly, a line of research is to attempt to mimic the development of bnAbs in natural infection
by immunising with a longitudinal sequence of trimers based on the Envs from an infected individual
who went on to generate bnAbs. An in-depth understanding of viral and antibody co-evolution
throughout the course of natural infection will no doubt prove invaluable in aiding the design of such immunisation regimens [132, 191, 213, 238, 239].

An alternative, albeit closely linked, approach is to design immunogens specifically targeted to initiate bnAb development. As noted previously, the development of bnAbs requires the activation of B cell lineages expressing gl-bnAbs, which are typically very poor at binding Env trimers [61-64]. This approach therefore involves the modification of immunogens in order to better engage gl-bnAbs [240], usually by the deletion of PNGS and variable loops to remove the steric occlusions which prevent gl-bnAb binding [241-243]. The laboratories of Schief and Nussenzweig have reported that BG505 SOSIP trimers containing, among other mutations, N133 and N137 PNGS deletions, are able to bind a germline-reverted version of the outer domain glycan-targeting bnAb, PGT121 [244].

Immunisation of mice engineered to express the predicted germline of PGT121 with this trimer, prior to boosting with a sequence of immunogens containing decreasing modifications, was able to elicit heterologous Tier-2 neutralising responses [245]. This provides proof of concept that immunisation with specifically designed immunogens can initiate bnAb development. There is, however, an apparent conflict between the need to delete PNGS sites in order to initiate bnAb development, and the recent observation that bnAbs were more likely to develop in individuals infected with isolates containing more intact glycan shields [76].

7. Conclusions

Antigenic mimicry is fundamental to most licensed vaccines. It is widely acknowledged that a protective HIV-1 vaccine will be based on the mimicry of the HIV-1 envelope spike, the sole target for bnAbs raised during infection. In addition to protein mimicry, Env-based immunogens will likely have to exhibit glycan mimicry, as many of the most potent bnAbs isolated to date recognise glycan structures within their epitopes. Thus, the characterisation of glycan epitopes, and an understanding of the principles governing their correct processing is needed in order to continue to guide the rational design of HIV-1 immunogens. Structural constraints imposed by the formation of native-like trimers restricts aspects of glycosylation processing, thus glycan analysis can help distinguish between native-like and non-native immunogens. However, native-like immunogens alone aren’t sufficient to induce bnAbs. As broadly neutralising epitopes are inherently immunooiquescent it is likely that additional strategies will be required in order to boost their immunogenicity.
Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank Ian Wilson, Andrew Ward, Bill Schief, Rich Wyatt, Jim Paulson, John Moore, Rogier Sanders, Robin Shattock and Raymond Dwek for their support and helpful discussions. This work is supported by the International AIDS Vaccine Initiative (M.C. and D.R.B), the Bill and Melinda Gates Foundation through the Collaboration for AIDS Discovery (OPP1084519 and OPP1115782, M.C. and D.B.R), the US National Institute of Allergy and Infectious Diseases (Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, UM1AI100663, D.R.B. and M.C.), and the Ragon Institute (D.R.B.). This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 681137 (K.J.D. and M.C.). K.J.D. also acknowledges the support of a Medical Research Council Career Development Fellowship.
References


neutralizing antibodies induced by native SOSIP.664 trimers that increase the arsenal of native-like Env immunogens. Proc Natl Acad Sci U S A. 2015;112:11947-52.


1273 2018;9:3693.
1277 Microheterogeneity at the PGT135 Antibody Recognition Site on HIV-1 gp120 Reveals a Molecular
1280 Deviation from Sigmoidal Neutralization Curves for HIV Broadly Neutralizing Monoclonal Antibodies.
1283 glycoproteome of virion-derived HIV-1 gp120 provides insights into broadly neutralizing antibody
1286 of Glycosylation Processing of a Glycan-Depleted Trimeric HIV-1 Immunogen Targeting Key B-Cell
1289 [181] Behrens AJ, Struwe WB, Crispin M. Glycosylation profiling to evaluate glycoprotein immunogens
1292 incorporation, and sensitivity to neutralizing antibodies of human immunodeficiency virus type 1
1294 1996;70:6431-6.
1295 [183] Pritchard LK, Spencer DI, Bonomelli C, Crispin M, Doores KJ. Cell- and Protein-Directed
1300 antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved
1303 basis for diverse N-glycan recognition by HIV-1-neutralizing V1-V2-directed antibody PG16. Nat Struct
1306 2018;140:5202-10.
1312 Envelope and Apex-Targeting Neutralizing Antibody Lineage Provides Benchmarks for Vaccine Design.
1313 [192] Wheeler SF, Rudd PM, Davis SJ, Dwek RA, Harvey DJ. Comparison of the N-linked glycans from


Dunlop DC. Antigenic mimicry as a strategy for AIDS vaccine design. [D.Phil.]2009.


