



# Glycosylation Benchmark Profile for HIV-1 Envelope Glycoprotein Production Based on Eleven Env Trimers

Eden P. Go,<sup>a</sup> Haitao Ding,<sup>b</sup> Shijian Zhang,<sup>c</sup> Rajesh P. Ringe,<sup>d</sup> Nathan Nicely,<sup>e</sup> David Hua,<sup>a</sup> Robert T. Steinbock,<sup>c</sup> Michael Golabek,<sup>d</sup> James Alin,<sup>e</sup> S. Munir Alam,<sup>e</sup> Albert Cupo,<sup>d</sup> Barton F. Haynes,<sup>e</sup> John C. Kappes,<sup>b,f</sup> John P. Moore,<sup>d</sup> Joseph G. Sodroski,<sup>c</sup> Heather Desaire<sup>a</sup>

Department of Chemistry, University of Kansas, Lawrence, Kansas, USA<sup>a</sup>; Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA<sup>b</sup>; Department of Cancer Immunology & Virology, Dana-Farber Cancer Institute, Department of Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts, USA<sup>c</sup>; Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, New York, USA<sup>d</sup>; Duke Human Vaccine Institute, Duke University Medical Center, Durham, North Carolina, USA<sup>e</sup>; Birmingham Veterans Affairs Medical Center, Research Service, Birmingham, Alabama, USA<sup>f</sup>

**ABSTRACT** HIV-1 envelope glycoprotein (Env) glycosylation is important because individual glycans are components of multiple broadly neutralizing antibody epitopes, while shielding other sites that might otherwise be immunogenic. The glycosylation on Env is influenced by a variety of factors, including the genotype of the protein, the cell line used for its expression, and the details of the construct design. Here, we used a mass spectrometry (MS)-based approach to map the complete glycosylation profile at every site in multiple HIV-1 Env trimers, accomplishing two goals. (i) We determined which glycosylation sites contain conserved glycan profiles across many trimeric Envs. (ii) We identified the variables that impact Env's glycosylation profile at sites with divergent glycosylation. Over half of the gp120 glycosylation sites on 11 different trimeric Envs have a conserved glycan profile, indicating that a native consensus glycosylation profile does indeed exist among trimers. We showed that some soluble gp120s and gp140s exhibit highly divergent glycosylation profiles compared to trimeric Env. We also assessed the impact of several variables on Env glycosylation: truncating the full-length Env; producing Env, instead of the more virologically relevant T lymphocytes, in CHO cells; and purifying Env with different chromatographic platforms, including nickel-nitrilotriacetic acid (Ni-NTA), 2G12, and PGT151 affinity. This report provides the first consensus glycosylation profile of Env trimers, which should serve as a useful benchmark for HIV-1 vaccine developers. This report also defines the sites where glycosylation may be impacted when Env trimers are truncated or produced in CHO cells.

**IMPORTANCE** A protective HIV-1 vaccine will likely include a recombinant version of the viral envelope glycoprotein (Env). Env is highly glycosylated, and yet vaccine developers have lacked guidance on how to assess whether their immunogens have optimal glycosylation. The following important questions are still unanswered. (i) What is the "target" glycosylation profile, when the goal is to generate a natively glycosylated protein? (ii) What variables exert the greatest influence on Env glycosylation? We identified numerous sites on Env where the glycosylation profile does not deviate in 11 different Env trimers, and we investigated the impact on the divergent glycosylation profiles of changing the genotype of the Env sequence, the construct design, the purification method, and the producer cell type. The data presented here give vaccine developers a "glycosylation target" for their immunogens, and they show how protein production variables can impact Env glycosylation.

**KEYWORDS** Env, glycopeptides, human immunodeficiency virus

The global AIDS pandemic is maintained by the occurrence of 2 to 3 million new HIV-1 infections annually (1). The development of prevention modalities, particularly a vaccine, that can interrupt these transmission events is a major priority for world health (2–4). Both passive immunization experiments in nonhuman primates (5–10) and correlates of protection in the RV144 clinical AIDS vaccine trial (11, 12) suggest that antibodies (Abs) may be important components of a protective immune response elicited by a vaccine.

The envelope glycoprotein (Env) spike, a heterodimeric trimer consisting of three gp120 exterior subunits and three gp41 transmembrane subunits (13, 14), is the only HIV-1-specific target for neutralizing antibodies (NAbs) (2, 4, 15). The appropriate presentation of Env to the immune system may be an essential part of a successful vaccine strategy. Unfortunately, the Env trimer has evolved features to minimize the elicitation and impact of neutralizing antibodies, including a heavy coat of carbohydrate (15–20). Approximately 50% of the mass of gp120 consists of carbohydrate, with most of the more than 24 potential *N*-linked glycosylation sites (PNGSs) utilized in most HIV-1 variants (15, 21–23). The gp41 component of the trimer typically contains 4 PNGSs (18, 24), and *O*-linked glycosylation sites also have been reported (25–27).

*N*-Linked glycans contribute to how HIV-1 evades NAbs (28). Because complex glycans resemble “self”-glycans and are generally less immunogenic than protein moieties, a significant fraction of the surface of the HIV-1 Env trimer is immunologically “silent” (20). Individual glycans can shield nearby epitopes directly, and the overall network of glycans contributes to the quaternary structure-dependent constraints on antibody access to the trimer (29, 30). Deleting or inserting specific *N*-linked glycans can render HIV-1 globally sensitive or resistant to antibodies that recognize different gp120 and gp41 epitopes (31, 32). Additionally, differences in Env immunogenicity have been observed when the glycosylation, but nothing else, changes on the immunogen (33, 34).

Many broadly neutralizing antibodies (bNAbs) generated in HIV-1-infected humans recognize relatively well-conserved gp120 epitopes that include glycan components. One group of such antibodies (2G12, PGT121, PGT128, and PGT135) binds high-mannose glycans (typically Man8 and Man9) on the gp120 outer domain that are conserved in a large fraction of HIV-1 strains (35, 36). All of these bNAbs except 2G12 also contact the gp120 protein surface, typically using long CDRH3 and CDRH1 loops to reach past the “glycan shield” (35, 37).

A second group of bNAbs recognizes glycan-related gp120 epitopes that are conserved in 75% to 80% of HIV-1 strains, sensitive to the quaternary structure of the HIV-1 Env trimer, and composed of the V1/V2 and V3 regions (the “trimer association domain”) of gp120 (38). These bNAbs (PG9, PG16, PGT145, CH01-04, and CAP256) have long, extended CDRH3 regions (39). Most of these bNAbs depend on an *N*-linked glycan at Asn 160, and some make contact with a second glycan at N156 (or N173 in some HIV-1 strains); Man3 or Man5 glycans at these positions are sufficient for antibody binding (38–40). A third group of bNAbs (35O22, PGT151, and PGT152) recognizes hybrid gp120-gp41 epitopes that include glycans (41, 42). In summary, because glycans are key components of various bNAb epitopes, the absence of the correct glycans from Env vaccine candidates is likely to reduce the chances of eliciting bNAbs that can recognize the known glycan-dependent epitopes on the virus.

Glycosylation analysis is a critical component of characterizing Envs, particularly those under consideration as immunogens for human trials. The first analysis of Env glycans on virions showed that predominantly high-mannose glycans were present and that those glycoforms were different from what typically had been observed on soluble, recombinant gp120 and gp140 proteins (43, 44). While these data were surprising and informative, the data set was incomplete because the glycans present at individual sites were not characterized. A subsequent attempt to analyze the glycans at each site on virion-derived Env yielded an incomplete data set—only 13 glycopeptides were identified in total, and most of the glycan sites were not detected (45). More recently, Go et al. published a site-by-site glycosylation analysis of a recombinant JR-FL (clade B) Env

trimer which yielded information on every site and identified >600 different glycoforms, a 30-fold increase over the earlier study on virion Env (25). The recombinant JR-FL trimer used in that study, a membrane-anchored gp150 protein purified by nickel-nitrilotriacetic acid (Ni-NTA) affinity, had multiple high-mannose glycans at specific sites, consistent with the original glycan-based virion analyses (43, 44). The major finding of the study by Go et al., that Env trimers contain multiple sites with high-mannose glycans, was replicated more recently in two analyses of the glycosylation pattern on the soluble, recombinant BG505 (clade A) SOSIP.664 gp140 trimer, purified with a 2G12 column (46, 47). In one BG505 analysis, data on 20 of 28 glycan sites were reported, with about 160 unique glycopeptides detected (46). Many of the characterized sites were exclusively occupied by high-mannose glycans (46). A more recent analysis of this same BG505 trimer provided glycosylation coverage at every site, with 230 glycopeptides reported (47). The studies on JR-FL gp150 and BG505 SOSIP.664 gp140 represent two examples of trimeric Env glycosylation, but the analyses were carried out using proteins based on different Env genotypes and on different construct designs and that were produced in different cell lines; neither study involved a natural HIV-1 target cell type or a full-length Env (25, 46, 47). Accordingly, it is unclear how these protein design and expression differences impacted the results. A subsequent analysis of glycosylation on a third genotype of Env (BAL, which is a clade B genotype), isolated from virions and purified by reverse-phase high-performance liquid chromatography (rHPLC), has been reported recently (48). Some of the glycan assignments in the latter virion study differed from ones identified in the earlier reports on the recombinant JR-FL and BG505 trimers for reasons that are not yet known. In summary, while multiple Env glycopeptide analyses all showed that trimeric Env is enriched for high-mannose glycans, it is unknown whether there is a consensus glycosylation profile at some sites, whether the producer cell or purification method significantly impacts the glycosylation profile, and whether the two truncated construct designs studied thus far—membrane-anchored gp150 and SOSIP-stabilized gp140—are good approximations of full-length Env.

We sought to elucidate as many of these unknowns about the glycosylation of Env trimers as possible in a single study. For the first time, we performed glycopeptide analyses on a large collection of Envs, by studying nine new proteins, in order to create a data set based on a total of 11 trimers. We then used the combined data set to explore the impact of multiple variables on Env glycosylation and to answer the following two key questions. (i) Is there a global “consensus” glycosylation profile of trimeric HIV-1 Env, present in a large set of diverse glycoproteins, that could be used for evaluating future Env immunogens? (ii) What variables contribute to Env glycosylation heterogeneity? The resulting identification of a native Env consensus glycosylation profile now provides vaccine developers with a “glycosylation target” for their immunogens, information that should assist in the eventual development of an effective HIV-1 vaccine.

## RESULTS AND DISCUSSION

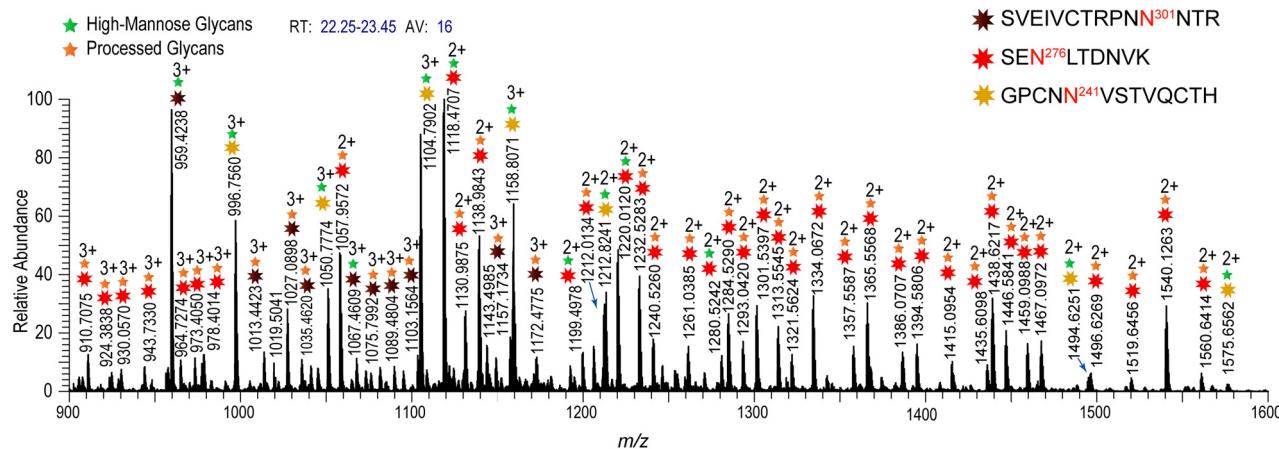
Here, we compare glycopeptide profiles for 11 different Env trimers and several other gp120 and gp140 proteins. Three of the trimer data sets are available in the literature, originally generated by the Desaire laboratory (25), the Crispin laboratory (46), and the Dell laboratory (48). Eight additional Env trimers were characterized for the present report. The full sequences and construct designs for these trimers are summarized in Fig. 1. We used the complete data set to identify a consensus glycosylation profile for trimers, which we then compared to the profiles for various gp120 and gp140 proteins. The glycosylation data on the latter proteins, many of which are former or current vaccine candidates, were mainly compiled from published reports, although the data on the gp140(-)FT protein from the C97ZA.012 genotype were obtained as part of the present study. Mass spectrometry data for this protein are shown in Fig. 2 and are used to demonstrate the analysis procedure performed on all the Envs.

The newly derived Env glycopeptide profiles were obtained using methodologies

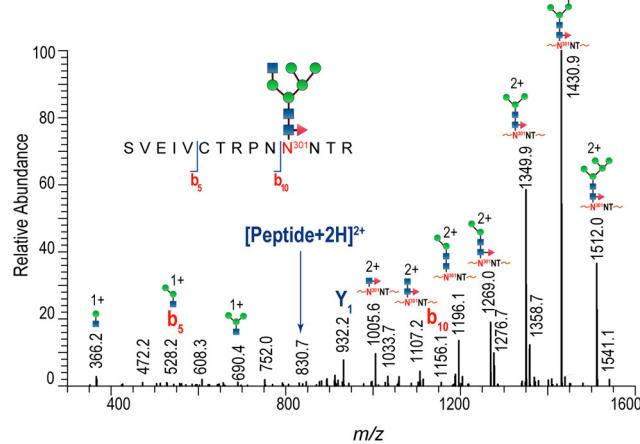
	Signal Peptide										C1	88
JR-FL_gp150 (-)Δ808	MPMGS1QP--LATLYLLG-----MLVASCLGVVEKLWTVVYGVWPWEATTTLFCASDAKAYDTEVHNWATHACVPTDPNFQEVLLENVTEHF											
JR-FL_gp160 (-)	MPMGS1QP--LATLYLLG-----MLVASVLAVEKLWTVVYGVWPWEATTTLFCASDAKAYDTEVHNWATHACVPTDPNFQEVLLENVTEHF											
BG505 gp140 SOSIP.664	--MDAMKRGGLCCVLLCAGAVFVSPSQEIHARFRRGARAENLWTVVYGVWPWEATTTLFCASDAKAYDTEVHNWATHACVPTDPNFQEVLLENVTEHF											
BG505 gp160	MPMGS1QP--LATLYLLG-----MLVASVLAENLWTVVYGVWPWEATTTLFCASDAKAYDTEVHNWATHACVPTDPNFQEVLLENVTEHF											
CH505 w4.3 gp150 (-)Δ808	MPMGS1QP--LATLYLLG-----MLVASCL---GMWTVVYGVWPWEAKTTLFCASDAKAYEKEVHNWATHACVPTDPNFQEMVLMKVNTENF											
CH505 w53.16 gp150 (-)Δ808	MPMGS1QP--LATLYLLG-----MLVASCL---GMWTVVYGVWPWEAKTTLFCASDAKAYEKEVHNWATHACVPTDPNFQEMVLMKVNTENF											
C97ZA.012 gp140 SOSIP.664	GARVG---NMWTVVYGVWPWEAKTTLFCASDAKAYEKEVHNWATHACVPTDPNFQEMVLMKVNTENF											
C97ZA.012 gp140-FT	-----AENLWVG---NMWTVVYGVWPWEAKTTLFCASDAKAYEKEVHNWATHACVPTDPNFQEMVLMKVNTENF											
135 138 141      V1 156 160      V2												
JR-FL_gp150 (-)Δ808	NMWKNNMVEQMQEDIISLWDQSLKPCVKLPLCVTLNCKDVNATNTND-SEG---TMRGEIKNCNFNIITTSIRDDEVKEYALFYKLDVVVPIDN-----											
JR-FL_gp160 (-)	NMWKNNMVEQMQEDIISLWDQSLKPCVKLPLCVTLNCKDVNATNTND-SEG---TMRGEIKNCNFNIITTSIRDDEVKEYALFYKLDVVVPIDN-----											
BG505 gp140 SOSIP.664	NMWKNNMVEQMHTDIIISLWDQSLKPCVKLPLCVTLNCTNVNN---ITDDMRGEKLNCNSCFNMTEELRDKKQKVYSLFYRLDVVQINENQGRNS											
BG505 gp160	NMWKNNMVEQMHTDIIISLWDQSLKPCVKLPLCVTLNCTNVNN---ITDDMRGEKLNCNSCFNMTEELRDKKQKVYSLFYRLDVVQINENQGRNS											
CH505 w4.3 gp150 (-)Δ808	NMWKNNDMVDQMHEDVISLWDQSLKPCVKLPLCVTLNCTNANATASNSIIEGMNNSIIEGMKNCNSFIITELRDKREKKNALFYKLDIVQLDG-----											
CH505 w53.16 gp150 (-)Δ808	NMWKNNDMVDQMHEDVISLWDQSLKPCVKLPLCVTLNCTNANATASNSIIEGMNNSIIEGMKNCNSFIITELRDKREKKNALFYKLDIVQLDG-----											
C97ZA.012 gp140 SOSIP.664	NMWKNNDMVDQMHEDIISLWDQSLKPCVKLPLCVTLHCTNATFKNN---VTNDMNKEIRNCNSFNTTEIRDKKQGYALFYRPDIVLLKEN---RN											
C97ZA.012 gp140-FT	NMWKNNDMVDQMHEDIISLWDQSLKPCVKLPLCVTLHCTNATFKNN---VTNDMNKEIRNCNSFNTTEIRDKKQGYALFYRPDIVLLKEN---RN											
186 187 197      230 C2 241      262      276												
JR-FL_gp150 (-)Δ808	--NNTSYRLISCDTSVITQACP KISFEPIPIHYCAPAGFAILKCNDFKGPCKNVSTVQCTHGIKPVYSTQLLLNGSLAEEVVIRSDNFTNNAKTI											
JR-FL_gp160 (-)	--NNTSYRLISCDTSVITQACP KISFEPIPIHYCAPAGFAILKCNDFKGPCKNVSTVQCTHGIKPVYSTQLLLNGSLAEEVVIRSDNFTNNAKTI											
BG505 gp140 SOSIP.664	NNSNKEYRLINCNTSAITQACP KVSFEPIPIHYCAPAGFAILKCKDKFFGTPCPSPVSTVQCTHGIKPVYSTQLLLNGSLAEEVVIRSENITNNAKNI											
BG505 gp160	NNSNKEYRLINCNTSAITQACP KVSFEPIPIHYCAPAGFAILKCKDKFFGTPCPSPVSTVQCTHGIKPVYSTQLLLNGSLAEEVVIRSENITNNAKNI											
CH505 w4.3 gp150 (-)Δ808	--NSSQYRLINCNTSAITQACP KVSFDPIPIHYCAPAGYAILKCNNTFTGTPCPNNSTVQCTHGIKPVYSTQLLLNGSLAEEVIIIRSENITNNVKTI											
CH505 w53.16 gp150 (-)Δ808	--NSSQYRLINCNTSAITQACP KVSFDPIPIHYCAPAGYAILKCNNTFTGTPCPNNSTVQCTHGIKPVYSTQLLLNGSLAEEVIIIRSENITDNVKTI											
C97ZA.012 gp140 SOSIP.664	NNSNSEYIILINCNASTITQACP KVNEDPIPIHYCAPAGYAILKCNNTFTSGKGPCKNVSTVQCTHGIKPVYSTQLLLNGSLAEKEIIIRSENLTNDVKTI											
C97ZA.012 gp140-FT	NNSNSEYIILINCNASTITQACP KVNEDPIPIHYCAPAGYAILKCNNTFTSGKGPCKNVSTVQCTHGIKPVYSTQLLLNGSLAEKEIIIRSENLTNDVKTI											
289 295 301      V3      332 339      C3 356 362												
JR-FL_gp150 (-)Δ808	IVQLKESEVINECTRPNNTTRKSISIHPGGRAFYTTGEIIDIGIRQAHCNISRAKWNDTLKQIVIKLREQFEN-KTIVFNHSSSGGDPEIVMHSFNCGGEFFYC											
JR-FL_gp160 (-)	IVQLKESEVINECTRPNNTTRKSISIHPGGRAFYTTGEIIDIGIRQAHCNISRAKWNDTLKQIVIKLREQFEN-KTIVFNHSSSGGDPEIVMHSFNCGGEFFYC											
BG505 gp140 SOSIP.664	IVQFNTPVQINCTRPNNTTRKSISIHPGQFYATGDIIDIRQAHCNVSASKATWETLGKVKQDLRKHFGNNTIIRFANSGGDLEVITTHSFNCGGEFFYC											
BG505 gp160	IVQFNTPVQINCTRPNNTTRKSISIHPGQFYATGDIIDIRQAHCNVSASKATWETLGKVKQDLRKHFGNNTIIRFANSGGDLEVITTHSFNCGGEFFYC											
CH505 w4.3 gp150 (-)Δ808	IVHLNESEVKIECTRPNNTTRKSISIHPGQFYATGQVIGDIREACYNIESKWNENTRQVSKKLKEYF-PHKNITFQPSSGGDLEITTHSFNCGGEFFYC											
CH505 w53.16 gp150 (-)Δ808	IVHLNESEVKIECTRPNNTTRKSISIHPGQFYATGQVIGDIREACYNIESKWNENTRQVSKKLKEYF-PHKNITFQPSSGGDLEITTHSFNCGGEFFYC											
C97ZA.012 gp140 SOSIP.664	IVHLNKSVEIVCTRPNNTTRKSISIHPGQFYATGQVIGDIREACYNIESKWNENTRQVSKKLKEYF-PHKNITFQPSSGGDLEITTHSFNCGGEFFYC											
C97ZA.012 gp140-FT	IVHLNKSVEIVCTRPNNTTRKSISIHPGQFYATGQVIGDIREACYNIESKWNENTRQVSKKLKEYF-PHKNITFQPSSGGDLEITTHSFNCGGEFFYC											
386 392 397 V4 406      C4 442 448      463 V5      C5												
JR-FL_gp150 (-)Δ808	NSTQLFNSTWNNN---TEGSNNTEG-NTITLP CRIKQIINMWQEVGKAMYAPP IRQIIRCSNSITGLLTRDGGINEN-GTEIFRPGGDMRDNRWSEL											
JR-FL_gp160 (-)	NSTQLFNSTWNNN---TEGSNNTEG-NTITLP CRIKQIINMWQEVGKAMYAPP IRQIIRCSNSITGLLTRDGGINEN-GTEIFRPGGDMRDNRWSEL											
BG505 gp140 SOSIP.664	NTSGLNFNTSWISNTS-VQGGSNSTGSNDSTLPCRIKQIINMWQRICQAMYAPP IRQIIRCSNSITGLLTRDGGSTNS-TTETFRPGGGDMRDNRWSEL											
BG505 gp160	NTSGLNFNTSWISNTS-VQGGSNSTGSNDSTLPCRIKQIINMWQRICQAMYAPP IRQIIRCSNSITGLLTRDGGSTNS-TTETFRPGGGDMRDNRWSEL											
CH505 w4.3 gp150 (-)Δ808	NTSSLNFNTRYMANSTDMANSTETNSTRTTIIHCRIKQIINMWQEVGRAMYAPPAGTGN TCI SNTGLLTRDGGKNNTETTFETFRPGGGDMRDNRWSEL											
CH505 w53.16 gp150 (-)Δ808	NTSSLNFNTRYMANSTDMANSTETNSTRTTIIHCRIKQIINMWQEVGRAMYAPPAGTGN TCI SNTGLLTRDGGKNNTETTFETFRPGGGDMRDNRWSEL											
C97ZA.012 gp140 SOSIP.664	NTTRLFNNNA-----TEDETITLP CRIKQIINMWQGVGRAMYAPPAGTGN TCKSNSITGLLLRDGGEDNK-TEELIFRPGGDMRDNRWSEL											
C97ZA.012 gp140-FT	NTTRLFNNNA-----TEDETITLP CRIKQIINMWQGVGRAMYAPPAGTGN TCKSNSITGLLLRDGGEDNK-TEELIFRPGGDMRDNRWSEL											
611 616 625      637      674												
JR-FL_gp150 (-)Δ808	KYKVVVKIEPLGVAPTKAKRVRVQSE-KSAVGIPLGVFLGLGAAGSTMGAASMTLTVQARL LLSGIVQQQNLLRAIEAQQRMLQLTWVGKIQLOARVLA											
JR-FL_gp160 (-)	KYKVVVKIEPLGVAPTKAKRVRVQSE-KSAVGIPLGVFLGLGAAGSTMGAASMTLTVQARL LLSGIVQQQNLLRAIEAQQRMLQLTWVGKIQLOARVLA											
BG505 gp140 SOSIP.664	KYKVVVKIEPLGVAPTRCKRRVVGRRRRRAVGIPLGVFLGLGAAGSTMGAASMTLTVQARL LLSGIVQQQNLLRAPEAQQHLLKLTWVGKIQLOARVLA											
BG505 gp160	KYKVVVKIEPLGVAPTRCKRRVVGRRRRRAVGIPLGVFLGLGAAGSTMGAASMTLTVQARL LLSGIVQQQNLLRAPEAQQHLLKLTWVGKIQLOARVLA											
CH505 w4.3 gp150 (-)Δ808	KYKVVVKIEPLGVAPTRCKRRVVGRRRRRAVGIPLGVFLGLGAAGSTMGAASMTLTVQARL LLSGIVQQQNLLRAPEAQQHLLKLTWVGKIQLOARVLA											
CH505 w53.16 gp150 (-)Δ808	KYKVVVKIEPLGVAPTRCKRRVVGRRRRRAVGIPLGVFLGLGAAGSTMGAASMTLTVQARL LLSGIVQQQNLLRAPEAQQHMLKLTWVGKIQLOARVLA											
C97ZA.012 gp140 SOSIP.664	KYKIELKPLGIAPT GCKRVRVVERRRRRRAVGIPLGVFLGLGAAGSTMGAASMTLTVQARL QLSSIVQQQNLLRAPEAQQHMLKLTWVGKIQLOTRVLA											
C97ZA.012 gp140-FT	KYKIELKPLGIAPT GCKRVRVVERERE-AVGIPLGVFLGLGAAGSTMGAASLTNLTVQARL QLSSIVQQQNLLRAPEAQQHMLKLTWVGKIQLOTRVLA											
KIFIMIVGGLVGLRLLVFTVLSIVNVRQGYSPLSFQTLLPAPRGPDPRPEGIEEEGERDRDRSGRLVNGFLALIWVDLRSCLCLFSYHRLRDLTTVTRIV												
JR-FL_gp160 (-)	KIFIMIVGGLVGLRLLVFTVLSIVNVRQGYSPLSFQTLLPAPRGPDPRPEGIEEEGERDRDRSGRLVNGFLALIWVDLRSCLCLFSYHRLRDLTTVTRIV											
BG505 gp140 SOSIP.664	KIFIMIVGGLIIGLRIVFAVL SIVHVRQGYSPLSFQTLLPAPRGPDPRPEGIEEEGERDRDRSGRLVNGFLALIWVDLRSCLCLFSYHRLRDLTTVTRIV											
BG505 gp160	KIFIMIVGGLIIGLRIVFAVL SIVHVRQGYSPLSFQTLLPAPRGPDPRPEGIEEEGERDRDRSGRLVNGFLALIWVDLRSCLCLFSYHRLRDLTTVTRIV											
CH505 w4.3 gp150 (-)Δ808	KIFIMIVGGLIIGLRIVFAVL SIVHVRQGYSPLSFQTLLPAPRGPDPRPEGIEEEGERDRDRSGRLVNGFLALIWVDLRSCLCLFSYHRLRDLTTVTRIV											
CH505 w53.16 gp150 (-)Δ808	KIFIMIVGGLIIGLRIVFAVL SIVHVRQGYSPLSFQTLLPAPRGPDPRPEGIEEEGERDRDRSGRLVNGFLALIWVDLRSCLCLFSYHRLRDLTTVTRIV											
C97ZA.012 gp140 SOSIP.664	KIFIMIVGGLIIGLRIVFAVL SIVHVRQGYSPLSFQTLLPAPRGPDPRPEGIEEEGERDRDRSGRLVNGFLALIWVDLRSCLCLFSYHRLRDLTTVTRIV											
C97ZA.012 gp140-FT	KIFIMIVGGLIIGLRIVFAVL SIVHVRQGYSPLSFQTLLPAPRGPDPRPEGIEEEGERDRDRSGRLVNGFLALIWVDLRSCLCLFSYHRLRDLTTVTRIV											
KSRIERGRGSGG-Y-----IP-EAPRDQAYVRKDGE-----EWVLLSTFL-												
JR-FL_gp150 (-)Δ808	ELLGR-----RGWEV LK YW NLL QY WS QEL GGG-----HHHHHH											
JR-FL_gp160 (-)	ELLGR-----RGWEV LK YW NLL QY WS QEL GGG-----HHHHHH											
BG505 gp140 SOSIP.664	ELLGR-----RGWEV LK YW NLL QY WS QEL GGG-----HHHHHH											
BG505 gp160	ELLGHSSLLKGRLRGWEGLKYLWNLNLLAYWGRELKISAINLFD TIAIAVAEWTDRVIEIGQRLCRAFLHIPRRIRQGLERAL-LVPRGSHHHHH											
CH505 w4.3 gp150 (-)Δ808	ELLGRSSLLKGRLRGWEGLKYLWNLNLLAYWGRELKISAINLFD TIAIAVAEWTDRVIEIGQRLCRAFLHIPRRIRQGLERAL-LVPRGSHHHHH											
CH505 w53.16 gp150 (-)Δ808	ELLGRSSLLKGRLRGWEGLKYLWNLNLLAYWGRELKISAINLFD TIAIAVAEWTDRVIEIGQRLCRAFLHIPRRIRQGLERAL-LVPRGSHHHHH											
C97ZA.012 gp140 SOSIP.664	ELLGRSSLLKGRLRGWEGLKYLWNLNLLAYWGRELKISAINLFD TIAIAVAEWTDRVIEIGQRLCRAFLHIPRRIRQGLERAL-LVPRGSHHHHH											
C97ZA.012 gp140-FT	ELLGRSSLLKGRLRGWEGLKYLWNLNLLAYWGRELKISAINLFD TIAIAVAEWTDRVIEIGQRLCRAFLHIPRRIRQGLERAL-LVPRGSHHHHH											

**FIG 1** Protein sequences of five HIV-1 Envs. Multiple-sequence alignment using Clustal Omega (67–69) was performed on the Envs used in this study. Conserved amino acid regions are colored in blue, and those less conserved are in black. The boundaries of the five conserved (C1 to C5, green lines) regions and the five variable (V1 to V5, purple lines) regions are shown, and potential N-linked glycosylation sites are colored in red. Sequence positions were standardized using the reference HIV-1 strain, HXB2.

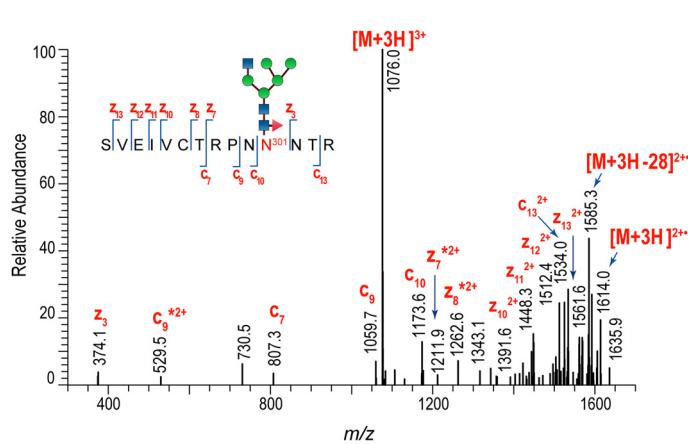
## A. Full MS



## B. CID



## C. ETD



**FIG 2** High-resolution and tandem MS data of C97ZA.012 gp140-FT. (A) Representative high-resolution MS data of a 293F cell-derived C97ZA.012 gp140-FT protein digest. The mass spectrum averaged within 22.25 to 23.45 min shows the elution of three glycopeptide species. Asterisks (\*) in the MS data denote glycopeptide peaks, and stars shaded in green and orange indicate the type of glycan. The glycan compositions of each observed glycopeptide peak in the mass spectra were verified from CID and ETD data. RT, retention time. (B) CID data of a glycopeptide with hybrid-type glycan and peptide portion, SVEIVCTRPN<sup>301</sup>NTR, showing the glycosidic cleavages. (C) ETD data of the same glycopeptide showing the peptide backbone cleavages.

that we developed over the past 10 years (25, 49–53). The various proteins were subjected to tryptic digestion, and the resulting glycopeptides were characterized using liquid chromatography-mass spectrometry (LC-MS) and tandem mass spectrometry (MS/MS). A typical high-resolution mass spectrum from a glycopeptide-rich region of one of the chromatograms, shown in Fig. 2A, contains a wide array of glycoforms that coelute. The glycopeptide compositions were identified by interrogating the high-resolution data along with the corresponding MS/MS data acquired on each of the relevant peaks. Both collision-induced dissociation (CID) and electron transfer dissociation (ETD) data were used to support the glycopeptide assignments reported here. Examples of MS/MS data of each type (i.e., CID and ETD) are shown in Fig. 2B and C. When every relevant peak in the extensive LC-MS data file was analyzed, as described previously (49, 50, 54), the outcome was a complete set of glycopeptide assignments for the glycoprotein. The full list of glycoforms detected for each of the proteins analyzed in this study is provided in Tables S1 through S8 in the supplemental material. Using the glycopeptide analysis methods described here, a highly rich glycosylation profile for each protein can be obtained. For example, ~600 unique glycopeptides were identified for the CHO cell-derived BG505 SOSIP.664 trimer.

Table 1 shows a comparison of the glycosylation profiles for the Env trimers at each commonly used glycosylation site. The predominant type of glycoform—either high-

**TABLE 1** Site-by-site comparison of the glycosylation profiles of 11 Env trimers

Sequence	gp120 Glycosylation Sites														gp41 Glycosylation Sites							Cell Line	Purification	
	88	133-139	156	160	187	197	230-234	241	262	276	289	295-301	332/334	339	356	386	389-448	463	611	616	625	637		
JR-FL gp160(-)	C	H	H	C	C			C	H	C		C	H	C	C	H	H	C	C	H	H	H	CHO	Ni-NTA/SEC
JR-FL gp150(-)Δ808 <sup>a</sup>	C	H	H	C	C			C	H	C		C	H	C	C	H	H	C	H	H	C	H	CHO	Ni-NTA <sup>d</sup>
JR-FL gp150(-)Δ808	H	H	H	H	C			C	H	C		H	H	H	C	H	H	C	H	H	C	H	M8166(T)	Ni-NTA/SEC
BG505 gp160	C	H	H	C	C	C	H		H	C		H	H	C	H	H	C	C	H	H	C	H	CHO	Ni-NTA/SEC
BG505 gp140 SOSIP.664	C	C	H	C	C	C	H		H	C		H	H	H	C	H	H	C	C	C	C	C	CHO	2G12/SEC
BG505 gp140 SOSIP.664 <sup>b</sup>	C	ND	H	H	C	C	H		H	H		H	H	H	C	H	H	C	C	C	ND	C	293T	2G12/SEC
CH505 w4.3 gp150(-)Δ808	C	H	H	C	C	C	H	H	H	C	H	H	H	H		H	H	C	H	H	H	H	CHO	2G12 <sup>d</sup>
CH505 w4.3 gp150(-)Δ808	C	H	H	H	C	C	H	H	H	C	H	H	H	H		H	H	C	H	H	H	H	M8166(T)	Ni-NTA/SEC
C97ZA.012 gp140 SOSIP.664	C	H	H	H	ND	C	H	H	H	C	H	H	H	C	C	H	H	C	H	H	C	C	293F	PGT151/SEC
CH505 w53 gp150(-)Δ808	C	H	H	H	C	C	H	H	H	H	H	C	H	H		H	H	C	H	H	H	H	CHO	2G12 <sup>d</sup>
BAL gp160 <sup>c</sup>	H	C	H	H	ND	C	ND	H	H	H	H	C	H	H	C	H	H	C	ND	ND	ND	ND	SUPT1-R5	rHPLC

**Key for Individual Sites:**

H = Mostly high-mannose glycans detected  
 C = Mostly processed glycans detected  
 ND=glycopeptides not detected at this site  
 (no entry) = glycosylation site not present for this protein

**Key for color coding:**

Consensus glycosylation profile across full Env set  
 Non-consensus due to genotype  
 Sites were glycosylation is influenced by Env construct design  
 Glycan type at this site can be influenced by producer cell type and genotype

<sup>a</sup>Analyzed previously. Data are from reference 25.<sup>b</sup>Analyzed previously. Data are from reference 46.<sup>c</sup>Analyzed previously. Data are from reference 48.<sup>d</sup>This protein was not SEC purified because it contained only trimers.

mannose glycan or processed glycan—is indicated at each site; these assignments are supported by data in the tables in the supplemental material, where every glycoform at every site is individually reported. Table 1 contains data from Envs of various genotypes (i.e., six different HIV-1 strains from multiple phylogenetic clades), construct designs (i.e., full-length gp160, truncated gp150, and soluble SOSIP.664 gp140), producer cell types (i.e., CHO, 293T, and M8166), and purification methods (i.e., Ni-NTA affinity, 2G12, and PGT151.) Detailed information about every Env in the study is provided in Table 1.

The most remarkable aspect of the data in Table 1 is the similarity among the Env trimers, even though the genotypes, construct designs, producer cells, and purification methods differ. More than half of the common glycosylation sites in the gp120 component of a trimeric Env have a conserved profile dominated by either high-mannose or processed glycoforms. The sites bearing predominately high-mannose glycans include N156, N262, N334, N389 to N448, and N463. The N389–N448 region of Env typically includes three to six glycosylation sites, which are all grouped together here for convenience. The sites in this region uniformly contain high-mannose glycans, despite the variation in the number of sites and their precise locations among different genotypes. Conversely, all the Envs have several sites where more highly processed glycoforms predominate. The latter sites include N187, N197, N230 to N234, N356, and N386. From here on, we refer to the sites that maintain a conserved glycosylation profile in each of the 11 Envs as the “consensus” profile.

At other gp120 sites, the majority glycan population is heterogenous, an observation that we now discuss starting with the ones with the least glycosylation heterogeneity. For N88, only one of the JR-FL Env trimers (gp150, expressed in M8166 cells) and BAL produced a glycan profile containing mostly high-mannose glycans; the other nine trimers in this study contained mostly complex glycoforms at this site. The M8166-expressed JR-FL gp150 also had a population of processed glycoforms at this site, and that population matched the majority glycoform type for all the other Env trimers reported in the tables in the supplemental material. For BAL, the Env glycopeptides were reported previously (48), but the glycopeptide coverage was low, with only four glycoforms detected at the N88 site. By comparison, we obtained between 9 and 30

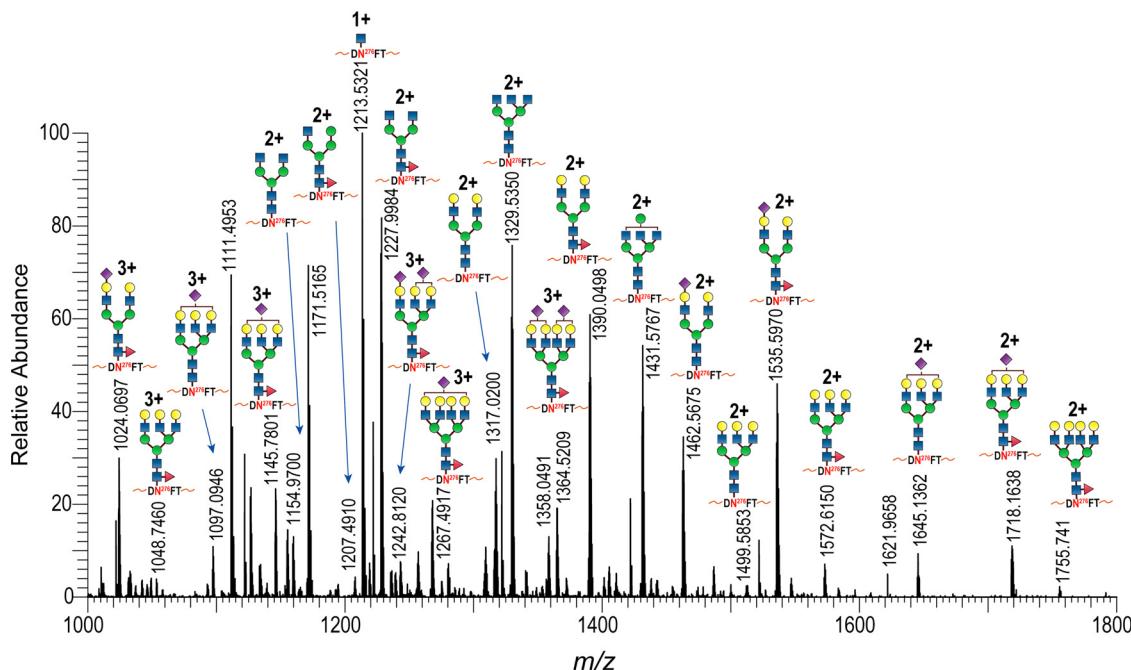
unique glycopeptides at this site for the various Env trimers analyzed here. One possible explanation for the divergent BAL data is that technical challenges associated with characterizing virion-derived Env might have resulted in reduced glycosylation coverage. If so, incomplete coverage might be responsible for the divergent glycosylation profile reported at N88 for this particular Env. Alternatively, BAL and JR-FL may represent atypical genotypes where high-mannose glycans predominate at N88 under certain conditions, such as when the Env is expressed in T cells.

For the N133-N139 glycosylation sites in the gp120 V1 variable region, only the BG505 SOSIP.664 trimer and the virion-derived BAL Env contain an outlier glycosylation profile in which complex glycans predominate. For the soluble BG505 trimer expressed in CHO cells, 54 glycopeptides were identified for the tryptic peptide containing N133 and N138; the vast majority (74%) of these were complex glycoforms. In contrast, only one glycopeptide was identified for virion-derived BAL, specifically, a complex glycoform. Given the substantial differences in the levels of glycosylation coverage, we defer consideration of the implications of the aberrant BAL glycosylation profile at this site until more complete data become available. Rather, we asked the following question. Why does the BG505 SOSIP.664 trimer contain mostly processed glycoforms at N133 and N138? When we instead expressed BG505 Env in the form of a full-length, membrane-anchored gp160, we again found that the glycosylation profile at N133 and N138 in CHO cells contained mostly high-mannose glycoforms, as was the case for the other Envs in the study (Table 1). This comparison between the SOSIP.664 gp140 and gp160 forms of BG505 Env suggests that the construct design difference contributes to the outlier glycosylation profile. More specifically, the greater processing of this V1 region glycan on the soluble SOSIP.664 trimer implies that these sites are more exposed during biosynthesis than they are on the membrane-anchored gp160. We also note that the region spanning glycosylation sites from N133 to N138 may also be influenced by the genotype of the SOSIP.664 construct; thus, high-mannose glycans are the majority glycoform present at this site on the C97ZA.012 SOSIP.664 protein that was also analyzed in this study.

Of the remaining sites on gp120, N241 contains the most conserved glycosylation profile. Thus, all the Env variants except those of the JR-FL and BG505 genotypes contain high-mannose glycans at this site. BG505 does not have a glycosylation site at N241, and all three JR-FL variants contain predominantly processed glycoforms. The unique glycosylation profile of JR-FL is clearly attributable to the genotype. It should be noted that the JR-FL genotype is unusual among HIV-1 strains in that it lacks a glycosylation site between N230 and N234. It is possible that the lack of a glycan at this site, by alleviating local steric constraints on the relevant enzymes, could have influenced how the nearby glycan at N241 is processed. Whether or not this explanation is correct, the data clearly indicate that how the N241 site is glycosylated is controlled by the genotype and not by the cell line, construct design, or purification method used.

In sharp contrast to N241, the glycosylation profile at N160 is clearly influenced by the producer cell type. In three different genotypes (JR-FL, BG505, and CH505 w4.3), the glycosylation profile at this site changes from mostly processed to mostly high-mannose glycoforms when the producer cell changes from CHO to T cells. For BG505, the cell line that produces protein with more high-mannose glycans is 293T; for the other two genotypes, the cell line is M8166, a CD4-positive T lymphocyte cell line. For some genotypes, T cells are not the only cells that produce immunogens with high-mannose glycans at this site; both the CH505 w53 Env trimer (produced in CHO cells) and the C97ZA.012 Env trimer (produced in 293F cells) contain high-mannose glycans as the majority glycoform.

Both the producer cell type and the Env genotype influence how the N295/N301 and N339 sites are glycosylated. These sites often contain high-mannose glycans, but not all of the 11 trimers contain high-mannose glycans as the major glycoform. Specifically, JR-FL trimers have mostly processed glycoforms at these sites, unless the protein is produced in M8166 cells. In addition, we also found that the C97ZA.012



**FIG 3** High-resolution MS data for N276 glycopeptides from JR-FL gp160 $(-)$ . Representative MS data illustrate that complex glycans are the predominant glycoforms present on the N276 glycosylation site in JR-FL gp160 $(-)$ . Prior to LC-MS analysis, the protein was treated with both Endo-H and trypsin, so all the high-mannose glycoforms combined into a single peak corresponding to the peptide plus one *N*-acetylhexosamine. This peak is detected at  $m/z$  1213.5321. Numerous other abundant peaks corresponding to complex glycopeptides from N276 populate the spectrum.

trimer is highly processed at N339 and likewise for CH505 w53 at N301. The CH505 w53 trimer has no glycan at N295.

The final site in gp120 whose glycosylation differed among the Env trimers in this study is N276. Here, the CH505 w53 gp150 is one example of a protein that contains mostly high-mannose glycans, and BG505 SOSIP.664 (from 293T cells) is another. In contrast, both BG505 SOSIP.664 (from CHO cells) and BG505 gp160 (also from CHO cells) contain mostly processed glycans at the N276 site. We interpret this finding to indicate that the glycosylation at this site can vary depending on the producer cell. The genotype is also a contributing factor to the glycosylation profile at N276. All three Envs of the JR-FL genotype contain processed glycoforms at this site. Figure 3 shows an example of a mass spectrum that includes many identifiable complex glycopeptides from the N276 glycosylation site. This spectrum was acquired while analyzing JR-FL gp160 $(-)$ .

The gp41 region is markedly more variably glycosylated across the data set of 11 Env trimers than the gp120 region. In particular, the N611 site is occupied by high-mannose glycans in some proteins but by complex glycoforms in others. For example, we found that the JR-FL gp160 $(-)$  and gp150 $(-)$  $\Delta$ 808 Envs were differentially glycosylated at this position, even though both proteins were expressed in the same CHO cell line and purified by Ni-NTA affinity chromatography. The construct design, therefore, must be driving force behind the variable glycosylation at this N611 site. We also note that both full-length gp160 proteins in the study (from two different genotypes) contained complex glycans at N611, while six of the eight truncated gp150 $(-)$  $\Delta$ 808 or gp140 $(-)$  proteins contained high-mannose glycans. Because construct design influences how the N611 site is glycosylated, at least for some genotypes, additional full-length (gp160) Envs of various genotypes will need to be analyzed to increase our database for this site. In summary, it is presently unknown whether native Env has a conserved profile of complex glycans at position N611.

The glycosylation differences seen at N616 are also driven by the construct design. The full-length BG505 gp160 protein contains  $>90\%$  high-mannose glycans at this site,

and yet both of the comparator BG505 SOSIP.664 trimers contain mostly processed glycoforms at N616. For the JR-FL genotype, the gp160(-) contains exclusively high-mannose glycans at this site, whereas the gp150(-)Δ808 comparator has mainly high-mannose glycans, although some complex glycoforms are also present. Taking the data together, both full-length Envs have more high-mannose glycans at N616 than their truncated comparator proteins. Thus, as for the N611 site, additional analyses of gp160 proteins are required to determine whether the glycosylation profile at N616 is conserved across all genotypes.

Glycosylation of the N625 site is highly variable in this data set. Two genotypes (CH505 w4.3 and CH505 w53) contain high-mannose glycans, one genotype (C97ZA.012) contains processed glycans as the main glycosylation type, and the glycosylation profiles of both JFRL and BG505 Envs vary depending on the construct design. For the two full-length gp160s, high-mannose glycans are the predominant glycoforms at N625, and Env truncation is responsible for changing the glycosylation profile at this site for both the JR-FL and BG505 genotypes. This data set shows that the glycosylation at N625 varies considerably when the construct length varies (i.e., gp140 or gp150 versus gp160); what is unknown at this time is whether full-length gp160 always contains a consensus glycosylation profile at N625 across diverse genotypes.

For N637, the only relevant influence is the genotype. Thus, all three Env constructs of the JR-FL genotype contain high-mannose glycans, as do both of the CH505 w4.3 variants and the CH505 w53 variant. In contrast, all three BG505 genotype proteins contain complex glycans at this site, as does C97ZA.012. Clearly, the genotype, and not the cell type or construct design, controls how the N637 site is glycosylated.

In summary, the data in Table 1 are color-coded to indicate those glycosylation sites that are part of a consensus glycosylation profile and those sites where glycosylation is variable. Additionally, the sites with variable glycosylation are categorized as sites that predominantly are influenced solely by the HIV-1 genotype or by both the genotype and producer cell type or instead by the Env construct design.

**Is the consensus glycosylation profile completely defined here?** The sites with a conserved glycosylation profile in every Env included in Table 1 are clearly part of the glycosylation consensus profile. Thus, our data indicate that, for these sites, the glycosylation consensus profile is independent of the purification method, the genotype, the cell type, or the construct design used to generate the trimers. The consensus glycosylation profile is, therefore, the intrinsic glycan profile that predominates on any properly folded trimeric Env, and hence it represents the native glycosylation profile. The consensus glycosylation profile (shown in green in Table 1) represents a target profile for vaccine developers striving to make natively glycosylated proteins.

To understand the native trimeric Env consensus profile more fully requires determining which additional sites, among those where we found that glycosylation varies, may also have a native consensus glycosylation profile on Env *in vivo*. In other words, it could be that the consensus profile is more extensive than the one that we have outlined above, but more studies are needed to determine if this is so. Below, we consider, independently, the impact of using proteins with truncated construct designs, proteins produced in different cell lines, and proteins purified by different methods. By doing so, we seek to answer the following question. Does the glycosylation variability shown in Table 1 represent native variability that arises from different Env genotypes, or was the variability introduced into the study via the use of, for example, truncated forms of Env?

**Impact of construct design.** In considering the impact of truncating the Env from its full-length gp160 form to a membrane-anchored gp150 or a soluble, SOSIP-stabilized gp140 form, we found that most of the glycosylation differences are located in the gp41 region of the proteins. Compared to the full-length BG505 gp160, the genotype-matched soluble SOSIP.664 gp140 protein has a different glycosylation profile at two of the four gp41 glycosylation sites (N616 and N625), whereas only the glycosylation at N133 to N139 in the gp120 region differs between these two con-

structs. A similar trend, of Env truncation impacting glycosylation in the gp41 region, is also observable in comparing the truncated gp150 and full-length gp160 proteins of the JR-FL genotype. Here, the predominant glycoforms in the gp120 regions of the two proteins are essentially unchanged, but the N625 and N611 glycans in gp41 do differ between the constructs. Additionally, more high-mannose glycans were detected at N616 for the full-length JR-FL gp160 than for its gp150 comparator. In summary, truncating Env constructs influences how gp41 is glycosylated. Additional studies are necessary to answer the following question. Are the glycans on N133, N611, N616, and N625 part of the consensus glycosylation profile for full-length gp160, which may be inappropriately glycosylated in truncated proteins? We note that these four sites may have a consensus profile, of high-mannose glycans at N133 to N139, N616, and N625 and complex glycans at N611, based on the data from the two full-length Envs in this study.

**Impact of producer cell type.** In considering the impact of the producer cell choice on Env glycosylation, we note that the N88, N160, N276, N295/N301, and N339 sites are where the glycosylation profile changes when proteins with the same genotype are produced in different cell lines. Some of these sites, i.e., N160, N295/N301, and N339, may contain a consensus profile of high-mannose glycans on native Env. Thus, for two different Env genotypes (i.e., CH505 and JR-FL), these sites were more likely to contain high-mannose glycans when the gp150 proteins were expressed in M8166 cells than when they were expressed in CHO cells. For N88 and N276, the glycosylation profiles were divergent between CH505 w4.3 and JR-FL gp150(-)Δ808 expressed in M8166 cells, so we do not expect that this site would contain a conserved glycosylation profile *in vivo*. Since M8166 cells are the most virologically relevant producer cell type used in this study—they are a CD4-positive T lymphocyte cell line and hence would be capable of being infected by an HIV-1 virus—additional studies of a more diverse panel of Envs produced in these cells are warranted to determine if N160, N295/N301, and N339 always have a conserved, native, high-mannose glycosylation profile when Envs are produced in CD4-positive T cells. In summary, we do not know at this time whether or not these sites are part of the consensus native Env glycosylation profile. Having stated that, the data in Table 1 indicate that the clear majority of Env glycosylation sites are not impacted by the choice of producer cell, at least with respect to whether their glycans are mostly in high-mannose form or are processed. An additional consideration, however, is that the cell type and, indeed, the construct design or the purification method may impact other aspects of glycan processing, such as the amount of sialylated or fucosylated glycans present on any given Env or at any given site. We include the full list of glycans at each glycosylation site for every newly analyzed protein in the supplemental material, and a preliminary comparison indicates that the M8166 cell line produces Envs with more sialylated glycoforms, for both CH505 w4.3 and JR-FL gp150(-)Δ808, than Envs of the same genotype produced in CHO cells. An in-depth study of the impact of the producer cell on the level of sialylation and fucosylation on Envs will be presented separately.

**Impact of genotype.** The Env genotype is solely responsible for the variability seen at two glycosylation sites, N241 and N637. While only one genotype (JR-FL) contains complex glycoforms at N241, all three Envs based on this genotype shared this unique feature. Similarly, the glycosylation profiles at N637 did not differ when the same genotype was expressed as a different construct design or in a different cell type or when the protein was purified differently; the implication is that the glycosylation profile at this site is fully controlled by the genotype. Additional studies will be needed to determine whether other sites, particularly N88, N133 to N139, N160, N276, N295 to N301, N339, N611, N616, and N625, can also vary in a genotype-dependent manner. Some of these sites should be investigated using full-length gp160 constructs, to eliminate any influences of Env truncation, and in some cases the Envs should be produced in the M8166 T cell line, for the reasons discussed above.

**Impact of purification method.** None of the variance in glycosylation reported in Table 1 is clearly attributable to the purification method used, although, in principle, how Envs are purified could affect their glycosylation profile. For example, it is possible that any differences observed between BG505 gp160 and SOSIP.664 gp140 could be influenced not just by the construct design but also by how they were purified: the gp160 initially via an Ni-NTA affinity column, the SOSIP.664 gp140 via a 2G12 affinity column. However, we think it is unlikely that the purification method had a major influence on the resulting glycosylation profiles, not least because the profiles of the gp160 and SOSIP.664 proteins were far more similar than different. One could argue that a 2G12-purified Env would be enriched for high-mannose glycans because its epitope involves a defined subset of such glycans. This was not, however, what we found. Thus, high-mannose glycans were actually more prevalent overall on the Ni-NTA affinity-purified gp160, while complex glycans were more prevalent at N133 to N138, N616, and N625 on the 2G12-purified SOSIP.664 gp140. Enrichment of protein containing high-mannose glycoforms on a 2G12 column would not be expected to lead to this outcome. We also note that when BG505 SOSIP.664 trimers, expressed in CHO cells, were affinity purified by different bNAbs (2G12, PGT151, and PGT145), there was no significant variation in their high-mannose contents (55).

A second pair of proteins was also purified by two different methods. CH505 w4.3 gp150(-)Δ808 were produced in CHO cells and M8166 cells and then purified by 2G12 and Ni-NTA affinity chromatography, respectively. The only glycosylation difference between these proteins was seen at N160, which is not involved in the 2G12 epitope. Furthermore, the 2G12-purified protein had fewer high-mannose glycans at N160 than its Ni-NTA affinity-purified counterpart. The glycosylation difference at this site is therefore more plausibly attributable to the producer cell type and not to the purification method. Additional evidence demonstrating a producer cell-dependent effect on glycosylation at N160 is described above. In summary, while different procedures were used to purify some of the proteins described in Table 1, the data argue against a meaningful influence of the Ni-NTA method versus the 2G12 affinity purification method on the resulting glycosylation variability.

One other point of variance in the Env purification methods was that three of the proteins were not purified by size exclusion chromatography (SEC); these proteins included JR-FL gp150(-)Δ808, CH505 w4.3 gp150(-)Δ808, and CH505 w53 gp150(-)Δ808. All three of these proteins were screened using size exclusion chromatography, and each of them produced only trimers. Since an SEC purification step would not improve the quality of these proteins, they were analyzed without performing this step. We do not expect that the lack of SEC purification of these three proteins impacted the study in any way.

Can a bNAb purification column have a major effect on the glycosylation profile? One protein that we studied here, the C97ZA.012 SOSIP.664 trimer, was purified by PGT151 affinity chromatography. PGT151 is a conformationally selective antibody, and a glycan is part of its epitope. This purification method was necessary to isolate appropriately folded C97ZA.012 trimers because a substantial subset of the nonselected gp140 proteins adopts multiple nonnative conformations (56). Compared to these irregularly folded proteins, the PGT151-purified C97ZA.012 trimers are enriched for high-mannose glycans (56). Therefore, in some cases the elimination of inappropriately assembled proteoforms with concomitantly aberrantly processed glycans is achievable using a bNAb column, as has been described previously elsewhere (56).

Does affinity purification performed with conformationally selective bNAb columns generate a more homogenous protein population solely by removing nonnative forms, or are functionally relevant forms also lost because they do not bind strongly to the bNAb column? Additional studies in this area are justified. In this context, Verkerke et al. have described an alternative way to purify SOSIP-stabilized gp140s that does not rely on any antibody-affinity columns (57). The use of this epitope-independent method, compared to a method using a bNAb selection column(s), could allow the

**TABLE 2** Comparing the consensus profile to glycosylation on gp120 and uncleaved gp140's<sup>a</sup>

<u>Sequence</u>	From	156	184-187	197	230	262	289	332/334	356	386-396	397-412	442	448	463
<b>Consensus</b>		HM	C	C	HM	HM	HM	HM	C	HM	HM	HM	HM	C
All 11 Envs	Table 1													
JR-FL gp140	Ref. 25													
C97ZA.012 gp140	Sup Data													
CON-S gp140	Ref. 58													
B.700010040.C9 gp140	Ref. 60													
C.1086 gp140	Ref. 60													
C.1086 gp120	Ref. 63													

Matches consensus   Does not match   Glycosylation site not present

<sup>a</sup>HM, high mannose; C, complex.

identification of any as-yet-unseen effects on the resulting glycosylation profile. While only a truly comparative experiment in which the sole variable was the purification method could yield a definitive answer, as of now we have found no evidence for a skewing effect on the glycosylation profile.

**Glycosylation of other vaccine candidates with truncated construct designs (gp120 and uncleaved gp140).** To demonstrate the profound differences in glycosylation that can be obtained when Env immunogens are expressed as a gp120 monomer or a gp140, we analyzed one additional uncleaved gp140(-) protein, of the C97ZA.012 genotype. We also compiled a list of glycosylation profiles from several previously analyzed gp120 and uncleaved gp140 proteins. We compared the glycosylation profiles of those proteins to the trimeric Env consensus profile determined here (Table 2). The JR-FL gp140(-) has no sites where high-mannose glycans predominate, and hence it deviates substantially from the consensus profile. The group M consensus Env, CON-S gp140, and a clade B transmitted/founder gp140, B.700010040.C9, have glycosylation profiles that are somewhat closer to the consensus profile for Env trimers but still differ from it at several sites. Finally, the clade C.1086 gp120, which will be tested in clinical trials (58), matches the consensus glycosylation profile at all sites except N156 and N397 to N412.

Why did every gp120 and gp140 tested here deviate from the glycosylation consensus of trimeric Env? On the basis of the comparative data in Table 2, we expect that trimeric Env immunogens may be necessary to obtain high-mannose glycans at N156 and N397 to N445, and yet the genotype, the culture and purification conditions, and the identity of the producer cell may all influence the extent to which other sites in the gp120 and gp140 proteins approximate the consensus profile. The glycosylation profile of the C.1086 gp120 protein was not highly divergent from the trimeric consensus, and we hypothesize that here the genotype may drive an atypically favorable glycosylation profile even without employing conformation-specific purification methods (e.g., bNAb columns). For the proteins with less-optimal glycosylation profiles than C.1086 gp120, some of the divergence is likely attributable to protein misfolding (59), which potentially can be overcome by using a bNAb column, or other methods, to select the most appropriately folded population. These various concepts could be tested experimentally in appropriately designed studies that take into account advances in construct design and Env purification procedures.

Finally, we note that the data in Tables 1 and 2, which can be used to show that uncleaved gp140 typically contains glycosylation profiles with an excess of processed glycoforms, are consistent with two prior studies addressing the impacts of furin cleavage and truncation on gp140 glycosylation. Go et al. had demonstrated that uncleaved, soluble gp140's of the JR-FL genotype contain predominantly

processed glycoforms at every glycosylation site, and yet membrane-anchored gp150's of the same genotype, which also lack a cleavage site, contain high-mannose glycans at most of the glycosylation sites (25). In that study, both sets of proteins were cleavage deficient, produced in the same cell type, and purified by the same methods; therefore, the presence of the membrane-anchoring region was the necessary feature that rescued Env's high-mannose glycans. Pritchard et al. have shown that for more-truncated Env's, such as soluble gp140's, a cleavage event is necessary to produce a protein with mainly high-mannose glycans (55). They showed that an uncleaved, SOSIP-stabilized BG505 gp140 contains only about 30% high-mannose glycans, while the cleaved version of the same protein contains about twice as many high-mannose glycoforms (55). Therefore, either the presence of a cleavage site (55) or the presence of a membrane-anchoring region (25) appears to be necessary to restore the high-mannose glycosylation profile associated with virions. The data in Tables 1 and 2 are consistent with both these prior studies; the uncleaved, soluble gp140s in Table 2 typically have aberrant glycosylation profiles, with too many processed glycoforms, and their glycosylation profiles are generally very different from those of all the cleaved, SOSIP-stabilized trimers and the uncleaved, membrane-anchored trimers in Table 1. Both these sets of trimers closely resemble those represented by the glycosylation data from virion-derived Env, as reported in reference 48 and displayed at the bottom of Table 1. We do not yet know why soluble gp140's require cleavage by furin to obtain glycosylation profiles with extensive high-mannose glycans, but the membrane-anchored gp150's and gp160's do not, so future studies on this topic are warranted. We suspect that the key driver of the high-mannose glycan profile is not cleavage *per se* but is instead the overall structure of the trimer, which can be driven by one device or another—including allowing cleavage to occur in the SOSIP trimer context or retaining in the construct the stabilizing elements present in the membrane-proximal external region (MPER), the transmembrane (TM), and—for the gp160's—the cytoplasmic domain.

In summary, we characterized a large panel of trimeric Envs in order to provide vaccine developers with a "glycosylation target" for future immunogens. In doing so, we mapped many consensus sites and identified four variables that may impact the glycosylation at nonconsensus sites. The consensus glycosylation target currently includes the following sites on gp120: N156, N187, N197, N230 to N234, N262, N289, N334, N356, N386, N389 to N448 (3 to 6 sites, depending on the genotype), and N463. Glycosylation profiling of these sites can guide vaccine developers in determining how to gauge the quality of glycosylation on their immunogen. Genotype differences contribute to Env glycosylation variability, certainly at N241 and N625 and probably also at N88 and N276. At other sites, there are likely to be genotype-dependent influences to factor in; here, the glycosylation profiles would probably be less useful for determining the quality of an immunogen. If additional data are obtained, some nonconsensus glycosylation sites that we have identified in this study may eventually be added as "consensus" sites.

We suggest three studies that would provide additional insight into the native Env glycosylation consensus. (i) A genetically diverse panel of full-length (gp160) Envs should be studied to understand whether or not there is a consensus glycosylation profile at the N133 to N139, N611, N616, and N625 sites. (ii) A genetically diverse panel of Envs produced in a highly relevant cell type, such as M8166 cells, should be studied, so that glycosylation at the N88, N160, N295/N301, and N339 sites can be further examined. (iii) Several pairwise comparisons should be made between the glycosylation profiles of Envs purified by affinity chromatography using a conformationally selective bNAb (e.g., PGT145 or PGT151 as appropriate to the genotype or construct design) versus an epitope-independent method (e.g., see reference 57), so that one can determine whether or not the bNAb column is biasing the glycan profile in Envs by enriching certain glycan types but not others.

We found that, compared to Env trimers, the glycosylation profiles of some gp120 monomers and gp140 proteins differed substantially from the consensus Env profile in some cases but less so in other cases. In the future, analysis of the glycosylation profile of candidate immunogens, particularly in comparison to the data we described here, will allow vaccine developers to judge the degree to which their immunogens mimic a trimeric consensus glycosylation profile. Because proper Env glycosylation is critically important to the binding of many broadly neutralizing antibodies, we expect that, all other things being equal, Envs with a native glycosylation profile will fare better as immunogens than those that lack this property.

In conclusion, we have defined the glycosylation consensus at many sites on gp120, identified a subset of sites where the glycosylation profile is clearly genotype specific, and proposed three future studies that would help to define glycosylation diversity at the nonconsensus sites more comprehensively. Understanding the native glycosylation profile of Env is important, and this report lays the foundation for researchers to assess and optimize how their Env immunogens are glycosylated. It will also be valuable to determine whether or not native glycosylation is necessary for Envs to induce a protective immune response. Experiments to assess whether and to what extent native glycosylation may improve Env immunogenicity have not been feasible to date, because critical knowledge was unavailable. Now that a benchmark profile that defines native glycosylation is emerging from our analyses, it will be possible to design appropriate studies in animal models.

## MATERIALS AND METHODS

**Expression, solubilization, and purification of membrane-anchored HIV-1 Env trimers.** For expression of membrane-anchored full-length HIV-1 BG505 gp160, full-length HIV-1 JR-FL gp160(-), and gp150(-)Δ808 gp150 from HIV-1 JR-FL, CH505 w4.3, and CH505 w59.16 genotypes, the *env* sequences were codon optimized and cloned into an HIV-1-based lentiviral vector. A heterologous signal sequence from CD5 replaced that of the wild-type HIV-1 Env. The proteolytic cleavage site between gp120 and gp41 is the wild-type sequence in the HIV-1 BG505 gp160; this glycoprotein also has a T332N change that restores a potential N-linked glycosylation site found in most HIV-1 Envs. The proteolytic cleavage site between gp120 and gp41 was altered in the other four membrane-anchored Envs, substituting serine residues for Arg 508 and Arg 511 (i.e., REKR to SEKS). The cytoplasmic tail of the gp150(-)Δ808 was truncated by replacement of the codon for Lys 808 with a sequence encoding (Gly)<sub>3</sub> (His)<sub>6</sub> followed immediately by a TAA stop codon. The full-length gp160 and gp160(-) glycoproteins have the sequence LVPRGSHHHHH at the carboxyl terminus. For controlling Env expression, the sequences encoding the gp160, gp160(-), and gp150(-)Δ808 glycoproteins were cloned immediately downstream of the tetracycline (Tet)-responsive element (TRE). The expression vector also included an internal ribosome entry site (IRES) and a contiguous puromycin (puro)-T2A-enhanced green fluorescent protein (EGFP) open reading frame, to allow puromycin resistance and EGFP expression, as reported previously (60).

The membrane-anchored full-length gp160 and gp160(-) glycoproteins were produced by expression in CHO cells. The membrane-anchored gp150(-)Δ808 glycoproteins were produced by expression either in CHO cells or in the M8166 T lymphocyte line. CHO-S cells (Invitrogen) and M8166 cells, modified to constitutively express the reverse Tet transactivator (rtTA), were transduced with Env-expressing lentivirus vectors pseudotyped with the vesicular stomatitis virus (VSV) G glycoprotein. Transduced CHO-S and M8166 cells were incubated in culture medium containing 1 μg/ml of doxycycline (dox) and then selected for 5 to 7 days in medium supplemented with 25 and 5 μg/ml, respectively, of puromycin. High-producer clonal CHO cell lines were derived using a FACSaria sorter (BD Biosciences) to isolate individual highly EGFP-fluorescent cells. The integrity of the recombinant *env* sequence in the clonal cell lines was confirmed by sequence analysis of PCR amplicons. Clonal cultures of CHO cells were adapted for growth in serum-free suspension culture medium (CDM4CHO; Thermo Fisher, Waltham, MA). For production of Env, CHO cells were expanded in suspension culture using a 14-liter New Brunswick BioFlo 310 fermentor (Eppendorf, Hauppauge, NY). Cultures were treated with 1 μg/ml doxycycline (Dox) after reaching a volume and density of 10 liters and >4 × 10<sup>6</sup> cells/ml, respectively. After 16 to 18 h of culture with Dox, the cells were harvested by centrifugation, snap-frozen in a dry ice-ethanol bath, and cryo-stored at -80°C until they were processed. M8166 cells expressing the gp150(-)Δ808 glycoproteins were adapted to culture in 1,050-cm<sup>3</sup> roller bottles (at ~9 rpm) using chemically defined medium (CDM4HEK293; Thermo Fisher, Waltham, MA) supplemented with 10% fetal bovine serum (FBS). For production of Env, the M8166 cells were expanded in roller bottles, and the cultures were treated with 1 μg/ml of Dox after reaching a volume and density of 500 ml and >4 × 10<sup>6</sup> cells/ml, respectively. After 48 h of culture with Dox, the cells were harvested and cryostored as described above for Env-expressing CHO cells.

**Extraction and purification of gp160, gp160(-), and gp150(-)Δ808 trimers.** With the exception of the CH505 gp150(-)Δ808 proteins expressed in CHO cells (see next section), all proteins were produced

in the following manner. Frozen cell pellets were homogenized in a homogenization buffer (250 mM sucrose, 10 mM Tris-HCl [pH 7.4], and a cocktail of protease inhibitors [Roche Complete tablets]). The membranes were extracted from the homogenates by differential centrifugation. The extracted crude membrane pellet was solubilized in a solubilization buffer containing 100 mM  $(\text{NH}_4)_2\text{SO}_4$ , 20 mM Tris-HCl (pH 8), 300 mM NaCl, 20 mM imidazole, 1% (wt/vol) Cymal-5 (Affymetrix), and a cocktail of protease inhibitors (Roche Complete tablets). The membranes were solubilized by incubation at 4°C for 1 h on a rocking platform. The suspension was centrifuged at 100,000 × g for 30 min. The supernatant was collected and mixed with a small volume of preequilibrated Ni-NTA beads (Qiagen) for 2 h on a rocking platform at 4°C. The mixture was then injected into a small column and washed with 30 bed volumes of a buffer containing 100 mM  $(\text{NH}_4)_2\text{SO}_4$ , 20 mM Tris-HCl (pH 8), 1 M NaCl, 30 mM imidazole, and 0.5% Cymal-5. The bead-filled column was eluted with a buffer containing 100 mM  $(\text{NH}_4)_2\text{SO}_4$ , 20 mM Tris-HCl (pH 7.4), 250 mM NaCl, 250 mM imidazole, and 0.5% (wt/vol) Cymal-5. The eluted gp160, gp160(-), or gp150(-)Δ808 glycoprotein solution was concentrated and then diluted in a sample buffer containing 20 mM Tris-HCl (pH 7.4), 250 mM NaCl, 100 mM  $(\text{NH}_4)_2\text{SO}_4$ , and 0.05% Cymal-6 for analysis of glycosylation profiles. For the gp150(-)Δ808 glycoproteins produced in M8166 cells, the concentrated eluate from the Ni-NTA beads was loaded onto a Yarra SEC-4000 column, using the sample buffer described above as the size exclusion chromatography (SEC) solvent. The fractions corresponding to the Env trimer peak were collected, pooled, concentrated, and then diluted in the sample buffer described above prior to glycosylation analysis.

For the CH505 gp150(-)Δ808 from CHO cells, the expressing cells ( $10 \times 10^6$ ) were frozen at -80°C and then thawed on ice. The cells were washed twice with 60 ml TNE buffer (25 mM Tris HCl [pH 7.5], 50 mM NaCl, 5 mM EDTA), with a centrifugation step performed in a swinging bucket centrifuge at 3,000 × g for 10 min after each wash. The cells were then suspended in 60 ml ice-cold TNE buffer, to which three protease inhibitor tablets had been added (Roche). Brij 58 was added to reach a 1% concentration, and then the mixture was incubated on ice for 1 h. The cells were homogenized with a tight Dounce homogenizer on ice and centrifuged for 1 h at 3,000 × g to remove nonsoluble material. The supernatant was collected, Cymal-5 was added to reach a 1% concentration, and the reaction mixture was incubated for 30 min with gentle shaking to produce the final lysate.

For the CH505 w4.3 gp150(-)Δ808 variant, the lysate was loaded onto a 10-ml lectin (*Galanthus nivalis*) column, washed to baseline with phosphate-buffered saline (PBS; 10 mM sodium phosphate [pH 7.2], 150 mM sodium chloride)-0.25% Cymal-5, and eluted with PBS-0.50% Cymal-5-0.5 M methyl α-D-mannopyranoside. Following elution, the protein was exchanged into PBS-0.25% Cymal-5 via four dilution/concentration cycles in Amicon Ultrafree 10-K molecular weight cutoff (MWCO) concentrators. The protein was then loaded onto a 10-ml column of 2G12 IgG coupled to Sepharose 4B (CNBr-activated Sepharose 4B; GE Life Sciences), following the manufacturer's protocol to couple IgG at a concentration of 10 mg/ml of media. The column was washed to baseline in PBS-0.25% Cymal-5, and then the protein was eluted in 10 mM Tris (pH 7.2)-3 M  $\text{MgCl}_2$ -0.25% Cymal-5. The eluted protein was buffer exchanged back into PBS-0.25% Cymal-5, concentrated, and quantified using the bicinchoninic acid (BCA) assay. An aliquot (1 mg) of the final sample was buffer exchanged into PBS-0.01% Cymal-6 for glycan analysis.

For CH505 w53.16 gp150(-)Δ808 purification, the lysate was loaded directly onto a 10-ml column of 2G12 IgG coupled to Sepharose 4B. All other steps of the purification were kept the same as described for CH505 w4.3 gp150(-)Δ808, except the final sample was kept in PBS-0.25% Cymal-5. Both the CH505 w4.3 and w53.16 gp150(-)Δ808 samples were screened on a Superdex 200 Increase column (GE Healthcare), and a single trimeric peak was present, without any monomers.

In addition to the proteins described above, several soluble gp140 trimers were studied. The BG505 SOSIP.664 construct was expressed in a stable CHO cell line, as described previously (59). The protein was purified first by using a 2G12-affinity column and then by SEC (59, 61, 62). The C97ZA.012 SOSIP.664 trimer and the C97ZA.012 gp140(-)FT were expressed in 293F cells by transient transfection, as described previously (56). The uncleaved C97ZA.012 gp140(-)FT, which contains a C-terminal His tag and a trimeric fibritin foldon at its C terminus, was first purified by Ni-NTA affinity chromatography, followed by SEC (56). The C97ZA.012 SOSIP.664 trimer was first purified via the use of a PGT151 bNAb affinity column and then by SEC (56). Note that the C97ZA.012 genotype descriptor, which was used in one earlier report (63) and is now used again here, is the same env genotype referred to elsewhere as CZA97.012 (56). Similarly, the gp140(-)FT descriptor is used here to indicate that this is an uncleaved soluble gp140 with a trimeric fibritin foldon domain, but exactly the same construct was previously referred to as gp140<sub>UNC</sub>-Fd-His (56).

**Partial deglycosylation of HIV-1 Env.** Samples containing 25 µg of HIV-1 Env were partially deglycosylated using Endo H. The pH of each sample was first adjusted to 5.5 before 2.5 µl of Endo H ( $\geq 5$  units/ml) was added. After mixing, the samples were incubated for 48 h at 37°C. The pH of the deglycosylated samples was adjusted to 8.0 prior to tryptic digestion, as described below.

**Proteolytic digestion of HIV-1 Env.** Two aliquots of Env samples (25 µg each) were added to a buffer containing 7 M urea-100 mM Tris (pH 8.0), reduced with 5 mM TCEP [tris(2-carboxyethyl) phosphine] at room temperature for 1 h, alkylated with 20 mM IAM (iodoacetamide) at room temperature for 1 h in the dark, and quenched the excess IAM with 20 mM DTT (dithiothreitol) for 15 min at room temperature. The reduced and alkylated samples were buffer exchanged and concentrated using a 50-K MWCO filter (Millipore) prior to protease digestion using trypsin alone or a combination of trypsin and chymotrypsin. All protease digestions were performed according to the suggested protocols of the manufacturers as follows. Digestion with trypsin was carried out with a 30:1 protein/enzyme ratio at 37°C for 18 h; digestion with chymotrypsin was carried out with a 25:1 protein/enzyme ratio at 37°C for 7 h; and when trypsin and chymotrypsin were used together in an overnight incubation at 37°C, the protein/enzyme ratio

was the same as that used for single-enzyme digestion. Following each digestion step, the reaction was quenched with acetic acid. The resulting Env digestion products were either analyzed immediately or stored at  $-20^{\circ}\text{C}$  until a later analysis. To ensure reproducibility of the method, the protein digestions were performed at least twice on different days with samples obtained from the same batch and analyzed using the same experimental procedure.

**Chromatography and mass spectrometry.** High-resolution LC-MS experiments were performed using an LTQ-Orbitrap Velos Pro (Thermo Scientific) mass spectrometer equipped with ETD (electron transfer dissociation) that was coupled to an Acquity ultra-high-performance liquid chromatography (UPLC) system (Waters). The mobile phases consisted of solvent A (99.9% deionized  $\text{H}_2\text{O}$ –0.1% formic acid) and solvent B (99.9%  $\text{CH}_3\text{CN}$ –0.1% formic acid). Five microliters of the sample ( $\sim 7 \mu\text{M}$ ) was injected onto a  $C_{18}$  PepMap 300 column (300  $\mu\text{m}$ ) at a flow rate of 5  $\mu\text{l}/\text{min}$ . A  $\text{CH}_3\text{CN}/\text{H}_2\text{O}$  multistep gradient was used that consisted of 3% solvent B for 5 min, followed a linear increase to 40% solvent B in 50 min and then a linear increase to 90% solvent B in 15 min. The column was held at 97% solvent B for 10 min before reequilibration. A short wash and blank run were performed between sample runs to eliminate any sample carryover. All mass spectrometric analyses were performed in a data-dependent mode as described below. The electrospray source was operated under the following conditions: source voltage of 3.0 kV, capillary temperature of  $250^{\circ}\text{C}$ , and S-lens radio frequency (RF) value between 45% and 55%. Data were collected in the positive ion mode. The data-dependent acquisition (DDA) mode was set up to sequentially and dynamically select the five most intense ions in the survey scan in the mass range of 400 to 2000  $m/z$  for alternating CID and ETD in the linear ion trap using a normalized collision energy of 30% for CID and an ion-ion reaction time of 100 to 150 ms for ETD. Full MS scans were measured at a resolution (R) of 30,000 at  $m/z$  400. Under these conditions, the measured R (full width at half-maximum [FWHM]) values in the Orbitrap mass analyzer were 20,000 at  $m/z$  1,000 and 17,000 at  $m/z$  1,500.

**Glycopeptide identification.** Raw data were analyzed using GlycoPep DB (54), GlycoPep ID (64), and GlycoMod (65). Details of the compositional analysis have been described previously (49–51, 66). Briefly, compositional analysis of glycopeptides with one glycosylation site was carried out by first identifying the peptide portion from tandem MS data. The peptide portion was inferred manually or by the use of GlycoPep DB from the  $Y_1$  ion, a glycosidic bond cleavage between the two *N*-acetylglucosamine at the pentasaccharide core. Once the peptide sequence was determined, plausible glycopeptide compositions were obtained using the high-resolution MS data and GlycoPep DB; the putative glycan candidate was then confirmed manually by identifying the  $Y_1$  ion and glycosidic cleavages from the CID data. Peptide fragment ions from ETD spectra of glycopeptides identified from a preceding CID scan were manually assessed for peptide fragment ions using Protein Prospector (<http://prospector.ucsf.edu>). Matched fragment ions with values within 0.5 Da of the theoretical value were accepted. For glycopeptides with multiple glycosylation sites, experimental masses of glycopeptide ions from the high-resolution MS data were converted to singly charged masses and submitted to GlycoMod. This program calculates plausible glycopeptide compositions from the set of experimental mass values entered by the user, compares these mass values with theoretical mass values, and then generates a list of plausible glycopeptide compositions within a specified mass error. Plausible glycopeptide compositions in GlycoMod were deduced by providing the mass of the singly charged glycopeptide ion, enzyme, protein sequence, cysteine modification, mass tolerance, and the possible types of glycans present in the glycopeptide. Plausible glycopeptide compositions obtained from the analysis were manually confirmed and validated from CID and ETD data.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/JVI.02428-16>.

**SUPPLEMENTAL FILE 1**, PDF file, 0.5 MB.

## ACKNOWLEDGMENTS

This study was supported by National Institutes of Health grants GM103547, AI094797, and AI125093 to Heather Desaire, P01AI110657 to John P. Moore, U19 AI067854 and GM095639 to John C. Kappes, and AI24755 to Joseph G. Sodroski and by Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery (CHAVI-ID) grant AI100645 to Barton F. Haynes. The work was also supported by the Virology, Genetic Sequencing, and Flow Cytometry Cores of the University of Alabama at Birmingham (UAB) Center for AIDS Research (NIH; P30 AI27767).

## REFERENCES

- UNAIDS. 2012. UNAIDS report on the global HIV/AIDS epidemic. United Nations Programme on HIV/AIDS, Geneva, Switzerland.
- Hoxie JA. 2010. Toward an antibody-based HIV-1 vaccine. Annu Rev Med 61:135–152. <https://doi.org/10.1146/annurev.med.60.042507.164323>.
- Klein E, Ho RJ. 2000. Challenges in the development of an effective HIV vaccine: current approaches and future directions. Clin Ther 22:295–314. [https://doi.org/10.1016/S0149-2918\(00\)80034-X](https://doi.org/10.1016/S0149-2918(00)80034-X).
- Letvin NL. 2005. Progress toward an HIV vaccine. Annu Rev Med 56: 213–223. <https://doi.org/10.1146/annurev.med.54.101601.152349>.
- Mascola JR, Lewis MG, Stiegler G, Harris D, VanCott TC, Hayes D, Louder

- MK, Brown CR, Sapan CV, Frankel SS, Lu Y, Robb ML, Katinger H, Birx DL. 1999. Protection of Macaques against pathogenic simian/human immunodeficiency virus 89.6PD by passive transfer of neutralizing antibodies. *J Virol* 73:4009–4018.
6. Shibata R, Igarashi T, Haigwood N, Buckler-White A, Ogert R, Ross W, Willey R, Cho MW, Martin MA. 1999. 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. <https://doi.org/10.1038/5568>.
  7. Hessell AJ, Hangartner L, Hunter M, Havenith CE, Beurskens FJ, Bakker JM, Lanigan CM, Landucci G, Forthal DN, Parren PW, Marx PA, Burton DR. 2007. Fc receptor but not complement binding is important in antibody protection against HIV. *Nature* 449:101–104. <https://doi.org/10.1038/nature06106>.
  8. Hessell AJ, Rakasz EG, Tehrani DM, Huber M, Weisgrau KL, Landucci G, Forthal DN, Koff WC, Poignard P, Watkins DL, Burton DR. 2010. 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. <https://doi.org/10.1128/JVI.01272-09>.
  9. Ruprecht RM. 2009. Passive immunization with human neutralizing monoclonal antibodies against HIV-1 in macaque models: experimental approaches. *Methods Mol Biol* 525:559–566, xiv. [https://doi.org/10.1007/978-1-59745-554-1\\_31](https://doi.org/10.1007/978-1-59745-554-1_31).
  10. Burton DR, Hessell AJ, Keele BF, Klasse PJ, Ketas TA, Moldt B, Dunlop DC, Poignard P, Doyle LA, Cavacini L, Veazey RS, Moore JP. 2011. Limited or no protection by weakly or nonneutralizing antibodies against vaginal SHIV challenge of macaques compared with a strongly neutralizing antibody. *Proc Natl Acad Sci U S A* 108:11181–11186. <https://doi.org/10.1073/pnas.1103012108>.
  11. Haynes BF, Gilbert PB, McElrath MJ, Zolla-Pazner S, Tomaras GD, Alam SM, Evans DT, Montefiori DC, Karnasuta C, Suthent R, Liao HX, DeVico AL, Lewis GK, Williams C, Pinter A, Fong Y, Janes H, DeCamp A, Huang Y, Rao M, Billings E, Karasavvas N, Robb ML, Ngauv V, de Souza MS, Paris R, Ferrari G, Bailer RT, Soderberg KA, Andrews C, Berman PW, Frahm N, De Rosa SC, Alpert MD, Yates NL, Shen X, Koup RA, Pitisutthithum P, Kaewkungwal J, Nitayaphan S, Rerks-Ngarm S, Michael NL, Kim JH. 2012. Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N Engl J Med* 366:1275–1286. <https://doi.org/10.1056/NEJMoa1113425>.
  12. Rerks-Ngarm S, Pitisutthithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, Premrri N, Namwat C, de Souza M, Adams E, Benenson M, Gurunathan S, Tartaglia J, McNeil JG, Francis DP, Stablein D, Birx DL, Chunsuttiwat S, Khamboonruang C, Thongcharoen P, Robb ML, Michael NL, Kunasol P, Kim JH. 2009. Vaccination with ALVAC and AIDSVA to prevent HIV-1 infection in Thailand. *N Engl J Med* 361:2209–2220. <https://doi.org/10.1056/NEJMoa0908492>.
  13. Robey WG, Safai B, Oroszlan S, Arthur LO, Gonda MA, Gallo RC, Fischeringer PJ. 1985. Characterization of envelope and core structural gene products of HTLV-III with sera from AIDS patients. *Science* 228:593–595. <https://doi.org/10.1126/science.2984774>.
  14. Allan JS, Coligan JE, Barin F, McLane MF, Sodroski JG, Rosen CA, Haseltine WA, Lee TH, Essex M. 1985. Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. *Science* 228: 1091–1094. <https://doi.org/10.1126/science.2986290>.
  15. Wyatt R, Sodroski J. 1998. The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* 280:1884–1888. <https://doi.org/10.1126/science.280.5371.1884>.
  16. Kwong PD, Doyle ML, Casper DJ, Cicala C, Leavitt SA, Majeed S, Steenbeke TD, Venturi M, Chaiken I, Fung M, Katinger H, Parren PW, Robinson J, Van RD, Wang L, Burton DR, Freire E, Wyatt R, Sodroski J, Hendrickson WA, Arthos J. 2002. HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* 420: 678–682. <https://doi.org/10.1038/nature01188>.
  17. McCaffrey RA, Saunders C, Hensel M, Stamatatos L. 2004. N-linked glycosylation of the V3 loop and the immunologically silent face of gp120 protects human immunodeficiency virus type 1 SF162 from neutralization by anti-gp120 and anti-gp41 antibodies. *J Virol* 78:3279–3295. <https://doi.org/10.1128/JVI.78.7.3279-3295.2004>.
  18. Reitter JN, Means RE, Desrosiers RC. 1998. A role for carbohydrates in immune evasion in AIDS. *Nat Med* 4:679–684. <https://doi.org/10.1038/nm0698-679>.
  19. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, Salazar-Gonzalez JF, Salazar MG, Kilby JM, Saag MS, Komarova NL, Nowak MA, Hahn BH, Kwong PD, Shaw GM. 2003. Antibody neutralization and escape by HIV-1. *Nature* 422:307–312. <https://doi.org/10.1038/nature01470>.
  20. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, Hendrickson WA, Sodroski JG. 1998. The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* 393:705–711. <https://doi.org/10.1038/31514>.
  21. Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ. 1990. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J Biol Chem* 265:10373–10382.
  22. Mizuuchi T, Spellman MW, Larkin M, Solomon J, Basa LJ, Feizi T. 1988. Carbohydrate structures of the human-immunodeficiency-virus (HIV) recombinant envelope glycoprotein gp120 produced in Chinese-hamster ovary cells. *Biochem J* 254:599–603. <https://doi.org/10.1042/bj2540599>.
  23. Zhu X, Borchers C, Bienstock RJ, Tomer KB. 2000. Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells. *Biochemistry* 39:11194–11204. <https://doi.org/10.1021/bi000432m>.
  24. Lee WR, Yu XF, Syu WJ, Essex M, Lee TH. 1992. Mutational analysis of conserved N-linked glycosylation sites of human immunodeficiency virus type 1 gp41. *J Virol* 66:1799–1803.
  25. Go EP, Herschhorn A, Gu C, Castillo-Menendez L, Zhang S, Mao Y, Chen H, Ding H, Wakefield JK, Hua D, Liao HX, Kappes JC, Sodroski J, Desaire H. 2015. Comparative analysis of the glycosylation profiles of membrane-anchored HIV-1 envelope glycoprotein trimers and soluble gp140. *J Virol* 89:8245–8257. <https://doi.org/10.1128/JVI.00628-15>.
  26. Stansell E, Canis K, Haslam SM, Dell A, Desrosiers RC. 2011. Simian immunodeficiency virus from the sooty mangabey and rhesus macaque is modified with O-linked carbohydrate. *J Virol* 85:582–595. <https://doi.org/10.1128/JVI.01871-10>.
  27. Stansell E, Panico M, Canis K, Pang PC, Bouché L, Binet D, O'Connor MJ, Chertova E, Bess J, Lifson JD, Haslam SM, Morris HR, Desrosiers RC, Dell A. 2015. Gp120 on HIV-1 virions lacks O-linked carbohydrate. *PLoS One* 10:e0124784. <https://doi.org/10.1371/journal.pone.0124784>.
  28. Stewart-Jones G, Soto C, Lemmin T, Chuang GY, Druz A, Kong R, Thomas P, Wagh K, Zhou T, Behrens AJ, Bylund T, Choi C, Davison J, Georgiev I, Joyce M, Kwon Y, Pancera M, Taft J, Yang Y, Zhang B, Shivatara S, Shivatara V, Lee CC, Wu CY, Bevley C, Burton D, Koff W, Connors M, Crispin M, Baxa U, Korber B, Wong CH, Mascola J, Kwong P. 2016. Trimeric HIV-1-Env structures define glycan shields from clades A, B, and G. *Cell* 165:813–826. <https://doi.org/10.1016/j.cell.2016.04.010>.
  29. Koch M, Pancera M, Kwong PD, Kolchinsky P, Grundner C, Wang L, Hendrickson WA, Sodroski J, Wyatt R. 2003. Structure-based, targeted deglycosylation of HIV-1 gp120 and effects on neutralization sensitivity and antibody recognition. *Virology* 313:387–400. [https://doi.org/10.1016/S0042-6822\(03\)00294-0](https://doi.org/10.1016/S0042-6822(03)00294-0).
  30. Mao Y, Wang L, Gu C, Herschhorn A, Desormeaux A, Finzi A, Xiang SH, Sodroski JG. 2013. Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer. *Proc Natl Acad Sci U S A* 110:12438–12443. <https://doi.org/10.1073/pnas.1307382110>.
  31. Haim H, Strack B, Kassa A, Madani N, Wang L, Courter JR, Princiotti A, McGee K, Pacheco B, Seaman MS, Smith AB III, Sodroski J. 2011. Contribution of intrinsic reactivity of the HIV-1 envelope glycoproteins to CD4-independent infection and global inhibitor sensitivity. *PLoS Pathog* 7:e1002101. <https://doi.org/10.1371/journal.ppat.1002101>.
  32. Haim H, Salas I, McGee K, Eichelberger N, Winter E, Pacheco B, Sodroski J. 2013. Modeling virus- and antibody-specific factors to predict human immunodeficiency virus neutralization efficiency. *Cell Host Microbe* 14: 547–558. <https://doi.org/10.1016/j.chom.2013.10.006>.
  33. Kong L, Sheppard NC, Stewart-Jones GBE, Robson CL, Chen H, Xu X, Krashias G, Bonomelli C, Scanlan CN, Kwong PD, Jeffs SA, Jones IM, Sattentau QJ. 2010. Expression-system-dependent modulation of HIV-1 envelope glycoprotein antigenicity and immunogenicity. *J Mol Biol* 403:131–147. <https://doi.org/10.1016/j.jmb.2010.08.033>.
  34. Bolmstedt A, Sjolander S, Hansen JE, Akerblom L, Hemming A, Hu SL, Morein B, Olofsson S. 1996. Influence of N-linked glycans in V4-V5 region of human immunodeficiency virus type 1 glycoprotein gp160 on induction of a virus-neutralizing humoral response. *J Acquir Immune Defic Syndr Hum Retroviril* 12:213–220. <https://doi.org/10.1097/00042560-19960700-00001>.
  35. Kong L, Lee JH, Doores KJ, Murin CD, Julien JP, McBride R, Liu Y, Marozsan A, Cupo A, Klasse PJ, Hoffenberg S, Caulfield M, King CR, Hua Y, Le KM, Khayat R, Deller MC, Clayton T, Tien H, Feizi T, Sanders RW,

- Paulson JC, Moore JP, Stanfield RL, Burton DR, Ward AB, Wilson IA. 2013. Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. *Nat Struct Mol Biol* 20:796–803. <https://doi.org/10.1038/nsmb.2594>.
36. Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien JP, Wang SK, Ramos A, Chan-Hui PY, Moyle M, Mitcham JL, Hammond PW, Olsen OA, Phung P, Fling S, Wong CH, Phogat S, Wrin T, Simek MD, Koff WC, Wilson IA, Burton DR, Poignard P. 2011. Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* 477:466–470. <https://doi.org/10.1038/nature10373>.
37. Pejchal R, Doores KJ, Walker LM, Khayat R, Huang PS, Wang SK, Stanfield RL, Julien JP, Ramos A, Crispin M, Depetris R, Katpally U, Marozsan A, Cupo A, Maloveste S, Liu Y, McBride R, Ito Y, Sanders RW, Ogohara C, Paulson JC, Feizi T, Scanlan CN, Wong CH, Moore JP, Olson WC, Ward AB, Poignard P, Schief WR, Burton DR, Wilson IA. 2011. A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. *Science* 334:1097–1103. <https://doi.org/10.1126/science.1213256>.
38. Walker LM, Phogat SK, Chan-Hui PY, Wagner D, Phung P, Goss JL, Wrin T, Simek MD, Fling S, Mitcham JL, Lehrman JK, Priddy FH, Olsen OA, Frey SM, Hammond PW, Kaminsky S, Zamb T, Moyle M, Koff WC, Poignard P, Burton DR. 2009. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* 326:285–289. <https://doi.org/10.1126/science.1178746>.
39. McLellan JS, Pancera M, Carrico C, Gorman J, Julien JP, Khayat R, Louder R, Pejchal R, Sastry M, Dai K, O'Dell S, Patel N, Shahzad-Ul-Hussan S, Yang Y, Zhang B, Zhou T, Zhu J, Boyington JC, Chuang GY, Diwanji D, Georgiev I, Kwon YD, Lee D, Louder MK, Moquin S, Schmidt SD, Yang ZY, Bonsignori M, Crump JA, Kapiga SH, Sam NE, Haynes BF, Burton DR, Koff WC, Walker LM, Phogat S, Wyatt R, Orwonyo J, Wang LX, Arthos J, Bewley CA, Mascola JR, Nabel GJ, Schief WR, Ward AB, Wilson IA, Kwong PD. 2011. Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* 480:336–343. <https://doi.org/10.1038/nature10696>.
40. Alam SM, Dennison SM, Aussedat B, Vohra Y, Park PK, Fernandez-Tejada A, Stewart S, Jaeger FH, Anasti K, Blinn JH, Kepler TB, Bonsignori M, Liao HX, Sodroski JG, Danishefsky SJ, Haynes BF. 2013. Recognition of synthetic glycopeptides by HIV-1 broadly neutralizing antibodies and their unmutated ancestors. *Proc Natl Acad Sci U S A* 110:18214–18219. <https://doi.org/10.1073/pnas.1317855110>.
41. Falkowska E, Le KM, Ramos A, Doores KJ, Lee JH, Blattner C, Ramirez A, Derking R, van Gils MJ, Liang CH, McBride R, von Bredow B, Shivate SS, Wu CY, Chan-Hui PY, Liu Y, Feizi T, Zwick MB, Koff WC, Seaman MS, Swiderek K, Moore JP, Evans D, Paulson JC, Wong CH, Ward AB, Wilson IA, Sanders RW, Poignard P, Burton DR. 2014. Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers. *Immunity* 40:657–668. <https://doi.org/10.1016/j.jimmuni.2014.04.009>.
42. Huang J, Kang BH, Pancera M, Lee JH, Tong T, Feng Y, Imamichi H, Georgiev IS, Chuang GY, Druz A, Doria-Rose NA, Laub L, Sliepen K, van Gils MJ, de la Peña AT, Derking R, Klasse PJ, Migueles SA, Bailer RT, Alam M, Pugach P, Haynes BF, Wyatt RT, Sanders RW, Binley JM, Ward AB, Mascola JR, Kwong PD, Connors M. 2014. Broad and potent HIV-1 neutralization by a human antibody that binds the gp41-gp120 interface. *Nature* 515:138–142. <https://doi.org/10.1038/nature13601>.
43. Bonomelli C, Doores KJ, Dunlop DC, Thaney V, Dwek RA, Burton DR, Crispin M, Scanlan CN. 2011. The glycan shield of HIV is predominantly oligomannose independently of production system or viral clade. *PLoS One* 6:e23521. <https://doi.org/10.1371/journal.pone.0023521>.
44. Doores KJ, Bonomelli C, Harvey DJ, Vasiljevic S, Dwek RA, Burton DR, Crispin M, Scanlan CN. 2010. Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. *Proc Natl Acad Sci U S A* 107:13800–13805. <https://doi.org/10.1073/pnas.1006498107>.
45. Yang W, Shah P, Toghi ES, Yang S, Sun S, Ao M, Rubin A, Jackson JB, Zhang H. 2014. Glycoform analysis of recombinant and human immunodeficiency virus envelope protein gp120 via higher energy collisional dissociation and spectral-aligning strategy. *Anal Chem* 86:6959–6967. <https://doi.org/10.1021/ac500876p>.
46. Behrens AJ, Vasiljevic S, Pritchard LK, Harvey DJ, Andev RS, Krumm SA, Struwe WB, Cupo A, Kumar A, Zitzmann N, Seabright GE, Kramer HB, Spencer DL, Royle L, Lee JH, Klasse PJ, Burton DR, Wilson IA, Ward AB, Sanders RW, Moore JP, Doores KJ, Crispin M. 2016. Composition and antigenic effects of individual glycan sites of a trimeric HIV-1 envelope glycoprotein. *Cell Rep* 14:2695–2706. <https://doi.org/10.1016/j.celrep.2016.02.058>.
47. Behrens AJ, Harvey DJ, Milne E, Cupo A, Kumar A, Zitzmann N, Struwe WB, Moore JP, Crispin M. 2 November 2016. Molecular architecture of the cleavage-dependent mannose patch on a soluble HIV-1 envelope glycoprotein trimer. *J Virol* <https://doi.org/10.1128/JVI.01894-16>.
48. Panico M, Bouché L, Binet D, O'Connor MJ, Rahman D, Pang PC, Canis K, North SJ, Desrosiers RC, Chertova E, Keele BF, Bess JW, Jr, Lifson JD, Haslam SM, Dell A, Morris HR. 2016. Mapping the complete glycoproteome of virion-derived HIV-1 gp120 provides insights into broadly neutralizing antibody binding. *Sci Rep* 6:32956. <https://doi.org/10.1038/srep32956>.
49. Go EP, Irungu J, Zhang Y, Dalpathado DS, Liao HX, Sutherland LL, Alam SM, Haynes BF, Desaire H. 2008. Glycosylation site-specific analysis of HIV envelope proteins (JR-FL and CON-S) reveals major differences in glycosylation site occupancy, glycoform profiles, and antigenic epitopes' accessibility. *J Proteome Res* 7:1660–1674. <https://doi.org/10.1021/pr7006957>.
50. Go EP, Chang Q, Liao HX, Sutherland LL, Alam SM, Haynes BF, Desaire H. 2009. Glycosylation site-specific analysis of clade C HIV-1 envelope proteins. *J Proteome Res* 8:4231–4242. <https://doi.org/10.1021/pr9002728>.
51. Go EP, Hewawasam G, Liao HX, Chen H, Ping LH, Anderson JA, Hua DC, Haynes BF, Desaire H. 2011. Characterization of glycosylation profiles of HIV-1 transmitted/founder envelopes by mass spectrometry. *J Virol* 85:8270–8284. <https://doi.org/10.1128/JVI.05053-11>.
52. Go EP, Hewawasam GS, Ma BJ, Liao HX, Haynes BF, Desaire H. 2011. Methods development for analysis of partially deglycosylated proteins and application to an HIV envelope protein vaccine candidate. *Int J Mass Spectrom* 305:209–216. <https://doi.org/10.1016/j.ijms.2010.11.009>.
53. Go EP, Liao HX, Alam SM, Hua D, Haynes BF, Desaire H. 2013. Characterization of host-cell line specific glycosylation profiles of early transmitted/founder HIV-1 gp120 envelope proteins. *J Proteome Res* 12:1223–1234. <https://doi.org/10.1021/pr300870t>.
54. Go EP, Rebecchi KR, Dalpathado DS, Bandu ML, Zhang Y, Desaire H. 2007. GlycoPep DB: a tool for glycopeptide analysis using a "Smart Search". *Anal Chem* 79:1708–1713. <https://doi.org/10.1021/ac061548c>.
55. Pritchard LK, Vasiljevic S, Ozorowski G, Seabright GE, Cupo A, Ringe R, Kim HJ, Sanders RW, Doores KJ, Burton DR, Wilson IA, Ward AB, Moore JP, Crispin M. 2015. Structural constraints determine the glycosylation of HIV-1 envelope trimers. *Cell Rep* 11:1604–1613. <https://doi.org/10.1016/j.celrep.2015.05.017>.
56. Ringe RP, Yasmeen A, Ozorowski G, Go EP, Pritchard LK, Guttman M, Ketras TA, Cottrell CA, Wilson IA, Sanders RW, Cupo A, Crispin M, Lee KK, Desaire H, Ward AB, Klasse PJ, Moore JP. 2015. Influences on the design and purification of soluble, recombinant native-like HIV-1 envelope glycoprotein trimers. *J Virol* 89:12189–12210. <https://doi.org/10.1128/JVI.01768-15>.
57. Verkerke HP, Williams JA, Guttman M, Simonich CA, Liang Y, Filipavicius M, Hu S-L, Overbaugh J, Lee KK. 2016. Epitope-independent purification of native-like envelope trimers from diverse HIV-1 isolates. *J Virol* 90:9471–9482. <https://doi.org/10.1128/JVI.01351-16>.
58. Zambonelli C, Dey AK, Hilt S, Stephenson S, Go EP, Clark DF, Wininger M, Labranche C, Montefiori D, Liao HX, Swanstrom RL, Desaire H, Haynes BF, Carfi A, Barnett SW. 2016. Generation and characterization of a bivalent HIV-1 subtype C gp120 protein boost for proof-of-concept HIV vaccine efficacy trials in Southern Africa. *PLoS One* 11:e0157391. <https://doi.org/10.1371/journal.pone.0157391>.
59. Go EP, Cupo A, Ringe R, Pugach P, Moore JP, Desaire H. 2015. Native conformation and canonical disulfide bond formation are interlinked properties of HIV-1 Env glycoproteins. *J Virol* 90:2884–2894. <https://doi.org/10.1128/JVI.01953-15>.
60. Hildebrandt E, Ding H, Mulky A, Dai Q, Aleksandrov AA, Bajrami B, Diego PA, Wu X, Ray M, Naren AP, Riordan JR, Yao X, DeLucas LJ, Urbatsch IL, Kappes JC. 2015. A stable human-cell system overexpressing cystic fibrosis transmembrane conductance regulator recombinant protein at the cell surface. *Mol Biotechnol* 57:391–405. <https://doi.org/10.1007/s12033-014-9830-5>.
61. Ringe RP, Sanders RW, Yasmeen A, Kim HJ, Lee JH, Cupo A, Korzun J, Derking R, van Montfort T, Julien JP, Wilson IA, Klasse PJ, Ward AB, Moore JP. 2013. Cleavage strongly influences whether soluble HIV-1 envelope glycoprotein trimers adopt a native-like conformation. *Proc Natl Acad Sci U S A* 110:18256–18261. <https://doi.org/10.1073/pnas.1314351110>.
62. Sanders RW, Derking R, Cupo A, Julien JP, Yasmeen A, de Val N, Kim HJ, Blattner C, de la Peña AT, Korzun J, Golabek M, de Los Reyes K, Ketas TJ, van Gils MJ, King CR, Wilson IA, Ward AB, Klasse PJ, Moore JP. 2013. 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. <https://doi.org/10.1371/journal.ppat.1003618>.
- 63. Go EP, Hua D, Desaire H. 2014. 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* 13:4012–4027. <https://doi.org/10.1021/pr5003643>.
  - 64. Irungu J, Go EP, Dalpathado DS, Desaire H. 2007. Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. *Anal Chem* 79:3065–3074. <https://doi.org/10.1021/ac062100e>.
  - 65. Cooper CA, Gasteiger E, Packer NH. 2001. GlycoMod—a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* 1:340–349. [https://doi.org/10.1002/1615-9861\(200102\)1:2<340::AID-PROT340>3.0.CO;2-B](https://doi.org/10.1002/1615-9861(200102)1:2<340::AID-PROT340>3.0.CO;2-B).
  - 66. Irungu J, Go EP, Zhang Y, Dalpathado DS, Liao HX, Haynes BF, Desaire H. 2008. Comparison of HPLC/ESI-FTICR MS versus MALDI-TOF/TOF MS for glycopeptide analysis of a highly glycosylated HIV envelope glycoprotein. *J Am Soc Mass Spectrom* 19:1209–1220. <https://doi.org/10.1016/j.jasms.2008.05.010>.
  - 67. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. 2013. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res* 41:W597–W600. <https://doi.org/10.1093/nar/gkt376>.
  - 68. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539. <https://doi.org/10.1038/msb.2011.75>.
  - 69. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* 38:W695–W699. <https://doi.org/10.1093/nar/gkq313>.