



Published in final edited form as:

Anal Chem. 2021 October 12; 93(40): 13651–13657. doi:10.1021/acs.analchem.1c03059.

DeGlyPHER: an ultrasensitive method for analysis of viral spike N-glycoforms

Sabyasachi Baboo¹, Jolene K Diedrich¹, Salvador Martínez-Bartolomé^{†,1}, Xiaoning Wang^{†,1}, Torben Schiffner^{†,2,3,4}, Bettina Groschel^{2,3}, William R Schief^{2,3,4}, James C Paulson^{1,2}, John R Yates III¹

¹Department of Molecular Medicine, The Scripps Research Institute, La Jolla, California 92037, United States

²Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, California 92037, United States

³IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, California 92037, United States

⁴The Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts 02139, United States

Abstract

Corresponding Authors: Sabyasachi Baboo – Department of Molecular Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States; sbaboo@scripps.edu, **John R Yates III** – Department of Molecular Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States; jyates@scripps.edu.

[†]Salvador Martínez-Bartolomé – Yatiri Bio, San Diego, California 92131, United States

[†]Xiaoning Wang – Angal Biotechnology Co., Ltd, Life & Health Town Industrial Park, Suzhou, Jiangsu 215000, China

[†]Torben Schiffner – Institut für Wirkstoffentwicklung, Universität Leipzig, Leipzig 04103, Germany

Jolene K Diedrich – Department of Molecular Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States;

Salvador Martínez-Bartolomé – Department of Molecular Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States;

Xiaoning Wang – Department of Molecular Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States;

Torben Schiffner – Department of Immunology and Microbiology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States;

Bettina Groschel – Department of Immunology and Microbiology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States;

William R Schief – Department of Immunology and Microbiology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States;

James C Paulson – Department of Molecular Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States;

Author Contributions:

S.B. conceived the method. S.B., J.K.D. and X.W. designed the experiments. S.B. processed the samples and analyzed data. J.K.D. performed LC-MS/MS. S.M.B. created the GlycoMSQuant tool. T.S. and B.G. expressed and purified HIV Env trimers. W.R.S., J.C.P. and J.R.Y. supervised the project. S.B. wrote the paper with contribution from all authors. All authors have given approval to the final version of the manuscript.

Supporting Information

Additional figures – S1–S6 (PDF).

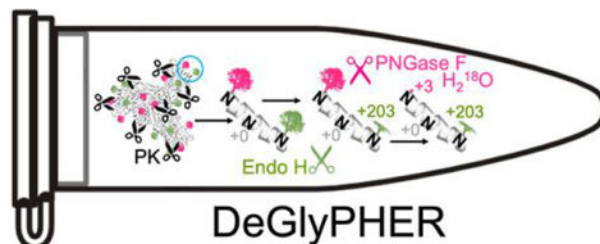
Code Availability:

GlycoMSQuant source code is freely available at <https://github.com/proteomicsyates/GlycoMSQuant> under a permissive Apache License 2.0.

The authors declare no competing interests.

Viruses can evade the host immune system by displaying numerous glycans on their surface “spike-proteins” that cover immune epitopes. We have developed an ultrasensitive “single pot” method to assess glycan occupancy and the extent of glycan processing from high-mannose to complex forms at each N-glycosylation site. Though aimed at characterizing glycosylation of viral spike-proteins as potential vaccines, this method is applicable for analysis of site-specific glycosylation of any glycoprotein.

Graphical Abstract



Keywords

proteomics; mass spectrometry; Proteinase K; N-glycan; heterogeneity; HIV; Env; viral spike; vaccine

INTRODUCTION:

Viral spike-proteins initiate virus entry into host cells and are the primary targets of vaccine design. Spike-proteins are often heavily N-glycosylated which help to shield the protein from the host immune response¹. These glycans add complexity to the production and characterization of recombinant protein-based vaccines². This is particularly a concern for characterization of the envelope spike-protein (Env) of the Human Immunodeficiency Virus (HIV) comprising a trimer with each monomer containing 26–30 unique N-linked glycosylation sites (NGS) as defined by the sequon NX[S/T], where X is any amino acid except P³. To address this analytical challenge several mass spectrometry-based strategies using multiple proteases^{4–6} have been implemented to create sufficient numbers of peptides unique to each glycosylation site^{3, 7, 8}. In these strategies, individual aliquots are digested with each protease and analyzed separately by liquid chromatography-mass spectrometry (LC-MS/MS), or pooled and analyzed together. To broadly characterize the nature of the glycosylation at each NGS, we had previously introduced the use of endoglycosidases that create residual mass signatures³. This helped us to determine the degree of glycan occupancy, and the degree of glycan processing – from the high mannose form that is initially attached to the protein, which may mature into the complex form when mannose residues are replaced by “terminal” monosaccharide sequences. To achieve coverage for all NGS, we had combined several digestions performed with different proteases and achieved >95% sequence coverage. Here we show that it is possible to replace these multiple proteolytic digestions with a single Proteinase K (PK) digestion, and moreover through careful choice of volatile buffers we have developed an improved “single pot” strategy with significantly increased sensitivity. We name this strategy to analyze glycoforms

as DeGlyPHER (*D*eglycosylation-dependent *G*lycan/*P*roteomic *H*eterogeneity *E*valuation Report).

EXPERIMENTAL SECTION:

Expression and purification of HIV Env trimers

BG505 SOSIP.664⁹ and BG505 SOSIP.664 MD39¹⁰ Env trimers were expressed and purified essentially as described previously¹⁰. Briefly, sequences with codons optimized for expression in human cells were synthesized and cloned into pHLSec between *AgeI*/*KpnI* by Genscript. The constructs were co-transfected with Furin-encoding plasmid, using polyethylenimine in Freestyle 293F cells cultured in 293 FreeStyle media (Thermo Fisher Scientific). Where indicated, 15 μ M sterile-filtered Kifunensine (Cayman Chemical) was added after transfection. After 6–7 days, supernatant was collected after passing through 0.22 μ m filter (Nalgene), and the C-terminally His-tagged trimers were purified using a HisTrap affinity column (Cytiva) with a linear elution gradient from 20–500 mM imidazole, followed by a Superdex 200 Increase SEC column (Cytiva) in Tris-buffered saline/TBS (20 mM Tris, 100 mM NaCl, pH 7.5). The oligomeric state and purity of trimer was verified (Figure S1) using size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS; DAWN HELEOS II/ Optilab T-rEX, Wyatt Technology). Total and protein molecular weights were determined using the protein-conjugate method in Astra 8.0.1.21 (Wyatt Technology).

Proteinase K treatment and deglycosylation

HIV Env trimer was exchanged to water using Microcon Ultracel-10 centrifugal device (Millipore Sigma). Trimer was reduced with 5 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl, Thermo Scientific) and alkylated with 10 mM 2-chloroacetamide (Sigma Aldrich) in 100 mM ammonium acetate for 20 min at room temperature (RT, 24°C). Initial protein-level deglycosylation (dispensable, as its omission did not change our results) was performed using 250 U of Endo H (New England Biolabs) for up to 5 μ g trimer, for 1 h at 37°C (pH 5.5–6.0). Trimer was digested with 1:25 Proteinase K (Sigma Aldrich) for 4 h at 37°C (pH 5.5–6.0). PK was denatured by incubating at 90°C for 15 min, then cooled to RT. Peptides were deglycosylated again with 250 U Endo H for 1 h at 37°C (pH 5.5–6.0), then frozen at –80°C and lyophilized. 100 U PNGase F (New England Biolabs) was lyophilized (for up to 5 μ g trimer), resuspended in 100 mM ammonium bicarbonate prepared in H₂¹⁸O (97% ¹⁸O, Sigma-Aldrich), and added to the lyophilized peptides. Reactions (5–20 μ l) were then incubated for 1 h at 37°C (pH 8.0–8.5), subsequently analyzed by LC-MS/MS. All reactions were performed in 0.2 ml PCR tubes on a thermocycler with heated lid.

Validating efficiency of glycosidases

BG505 SOSIP.664 HIV Env trimer glycosylated with only high-mannose N-glycans (purified from cells treated with Kifunensine¹¹, which inhibits processing of high-mannose N-glycans to complex N-glycans during protein maturation), was sequentially treated with Endo H, followed by PNGase F. After treatment with both enzymes, 99.2% of identified peptides were N+203; 100% of peptides identified with only PNGase F treatment were N+3 (Figures S2a,b; proportions do not consider unoccupied NGS because they remained similar

in both experiments – 8–10%). We realize the possibility that glycosidase PNGase F may occasionally cleave the remnant GlcNAc (N-Acetylglucosamine)¹² post-Endo H processing of high-mannose/hybrid N-glycans and thus, convert the mass modification characteristic of high-mannose/hybrid (+203) to complex (+3) N-glycans, which would affect our analyses. However, we have not observed any significant evidence of this possibility in our results, though it may explain why we observe a few peptides with +3 mass modified NGS in Kifunensine treated samples (Figure S2a).

Trypsin proteolysis

The Proteinase K/deglycosylation method described above was followed, except PK was replaced with trypsin and reactions were incubated overnight at 37°C. Trypsin generated a lower total number of peptides than PK, but we obtained >95% sequence coverage, including 26 of 27 NGS (Figure S3a). Variations in N-glycan microheterogeneity at certain NGS may be explained by low sampling at these sites (N386, N392) or difference in cleavage-specificity between PK and trypsin (N88, N611) (Figures S3b,c).

LC-MS/MS

Q Exactive HF-X with EASY-nLC 1200—Samples were analyzed on an Q Exactive HF-X mass spectrometer (Thermo). Samples were injected directly onto a 25 cm, 100 μ m ID column packed with BEH 1.7 μ m C18 resin (Waters). Samples were separated at a flow rate of 300 nL/min on an EASY-nLC 1200 (Thermo). Buffers A and B were 0.1% formic acid in 5% and 80% acetonitrile, respectively. The following gradient was used: 1–25% B over 160 min, an increase to 40% B over 40 min, an increase to 90% B over another 10 min and 30 min at 90% B for a total run time of 240 min. Column was re-equilibrated with solution A prior to the injection of sample. Peptides were eluted from the tip of the column and nanosprayed directly into the mass spectrometer by application of 2.8 kV at the back of the column. The mass spectrometer was operated in a data dependent mode. Full MS1 scans were collected in the Orbitrap at 120,000 resolution. The ten most abundant ions per scan were selected for HCD MS/MS at 25 NCE. Dynamic exclusion was enabled with exclusion duration of 10 s and singly charged ions were excluded.

timsTOF Pro with Evosep—One Samples were loaded onto EvoTips following manufacturer protocol. The samples were run on an Evosep One (Evosep) coupled to a timsTOF Pro (Bruker Daltonics). Samples were separated on a 15 cm \times 150 μ m ID column with BEH 1.7 μ m C18 beads (Waters) and integrated tip pulled in-house using either the 30 SPD or 15 SPD methods. Mobile phases A and B were 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively. MS data was acquired in PASEF mode with 1 MS1 survey TIMS-MS and 10 PASEF MS/MS scans acquired per 1.1 s acquisition cycle. Ion accumulation and ramp time in the dual TIMS analyzer was set to 100 ms each and we analyzed the ion mobility range from $1/K_0 = 0.6$ Vs cm^{-2} to 1.6 Vs cm^{-2} . Precursor ions for MS/MS analysis were isolated with a 2 Th window for $m/z < 700$ and 3 Th for $m/z > 700$ with a total m/z range of 100–1700. The collision energy was lowered linearly as a function of increasing mobility starting from 59 eV at $1/K_0 = 1.6$ VS cm^{-2} to 20 eV at $1/K_0 = 0.6$ Vs cm^{-2} . Singly charged precursor ions were excluded with a polygon filter, precursors for

MS/MS were picked at an intensity threshold of 2,500, target value of 20,000 and with an active exclusion of 24 s.

Data Processing

Protein and peptide identification were done with Integrated Proteomics Pipeline (IP2, Bruker Scientific LLC). Tandem mass spectra were extracted from raw files using RawConverter¹³ (timstofCoverter for timsTOF Pro data) and searched with ProLuCID¹⁴ against a database comprising UniProt reviewed (Swiss-Prot) proteome for *Homo sapiens* (UP000005640), UniProt amino acid sequences for Endo H (P04067), PNGase F (Q9XBM8), and Proteinase K (P06873), amino acid sequences for BG505 SOSIP.664⁹ and BG505 SOSIP.664 MD39¹⁰ (including a preceding secretory signal sequence and followed by 6xHis-tag), and a list of general protein contaminants. The search space included no cleavage-specificity (all fully tryptic and semi-tryptic peptide candidates for trypsin treatment). Carbamidomethylation (+57.02146 C) was considered a static modification. Deamidation in presence of H₂¹⁸O (+2.988261 N), GlcNAc (+203.079373 N), oxidation (+15.994915 M) and N-terminal pyroglutamate formation (−17.026549 Q) were considered differential modifications. Data was searched with 50 ppm precursor ion tolerance and 50 ppm fragment ion tolerance. Identified proteins were filtered using DTASelect2¹⁵ and utilizing a target-decoy database search strategy to limit the false discovery rate to 1%, at the spectrum level¹⁶. A minimum of 1 peptide per protein and no tryptic end (or 1 tryptic end when treated with trypsin) per peptide were required and precursor delta mass cut-off was fixed at 10 ppm for data acquired with Q Exactive HF-X or 20 ppm for data acquired with timsTOF Pro. Statistical models for peptide mass modification (modstat) were applied (trypstat was additionally applied for trypsin-treated samples). Census2¹⁷ label-free analysis was performed based on the precursor peak area, with a 10 ppm precursor mass tolerance and 0.1 min retention time tolerance. “Match between runs” was used to find missing peptides between runs for Q Exactive HF-X data (for timsTOF Pro data, reconstructed-MS1 based chromatograms combining isotope peaks for all triggered precursor ions were pre-generated, and then chromatograms were assigned to identified peptides for quantitative analysis, without retrieving missing peptides).

Data Analysis using GlycoMSQuant

Our new tool (v.1.4.1, <https://github.com/proteomicsyates/GlycoMSQuant>) was implemented to automate the analysis and to visualize the results. GlycoMSQuant summed precursor peak areas across replicates, discarded peptides without NGS, discarded misidentified peptides when N-glycan remnant-mass modifications were localized to non-NGS asparagines and corrected/fixed N-glycan mislocalization where appropriate. The results were aligned to NGS in Env of HXB2¹⁸ HIV-1 variant.

Precursor peak area was calculated by Census2¹⁷ from extracted-ion chromatogram (XIC) for each peptide in each replicate. For each NGS (NX[S]T), where X is any amino acid except P, the “N-glycosylation state” represented by proportions for unoccupied (+0, *u*), complex (+2.988261, *c*) and, high-mannose/hybrid (+203.079373, *h*) N-glycans was calculated as follows.

The sum of the precursor peak areas $S_{g,pepz}$ was calculated as:

$$S_{g,pepz} = \sum xic_{pepz}$$

where N-glycosylated peptides with the same sequences and charge were grouped together ($pepz$), g is the N-glycosylation state $\in G(u, c, h)$, and xic is the precursor peak area.

For each group ($pepz$), the abundance proportion $\%_{g,pepz}$ of each N-glycosylation state $g \in G$ was calculated as:

$$\%_{g,pepz} = \frac{S_{g,pepz}}{\sum_{i \in G} S_{i,pepz}}$$

Finally, as each NGS may be covered by multiple groups ($pepz$), the proportion of each N-glycosylation state g for a particular NGS (ngs) is calculated as the mean of all proportions $\%_{g,pepz}$ of all groups ($pepz$) covering this NGS:

$$\%_{ngs,g} = \frac{1}{n_{ngs}} \sum \%_{g,pepz}$$

where n_{ngs} is the number of groups ($pepz$) covering a particular NGS.

The standard error of mean of the proportion of each N-glycosylation state $g \in G$ for a particular NGS ($SEM_{ngs,g}$) was calculated as:

$$SEM_{ngs,g} = \frac{s_{ngs,g}}{\sqrt{n_{ngs}}}$$

where, $s_{ngs,g}$ is the standard deviation of $\%_{g,pepz}$ from all groups ($pepz$) covering a particular NGS.

Pairwise statistical comparisons of experiments (a and b) were performed for each $g \in G$ at each NGS using proportion values $\%_{g,pepz}$ of groups ($pepz$) sharing the NGS, applying the Mann-Whitney U test¹⁹. Testing $\%_{g,pepz,a}$ vs. $\%_{g,pepz,b}$ individually for u , c and h at each NGS, we calculated p -values that were subjected to multiple hypothesis correction using the Benjamini-Hochberg (BH) method²⁰. If the corrected p -value was <0.05 (for u , c or h at any NGS), then the difference was considered statistically significant.

For Figure S4, published data from our previous approach³ was reanalyzed using the data analysis workflow described here. Briefly, the data analyzed is from 3 replicates of 3 conditions, each separately analyzed by LC-MS/MS, with total starting material of 90 μ g protein, and Census2¹⁷ label-free analysis was performed simultaneously on all 9 experiments without “match between runs”, and the results analyzed by GlycoMSQuant. This was compared with data obtained from a single LC-MS/MS run (QE-HFX mass-spectrometer coupled to EASY-nLC 1200 HPLC or QE-HFX/nLC) with 0.5 μ g of peptides generated starting with 5 μ g purified protein.

RESULTS AND DISCUSSION:

PK is a broadly specific serine-protease that has previously been exploited for its potential to generate overlapping peptides and high sequence coverage²¹. The redundancy afforded by overlapping sequences significantly increases confidence in identifications, especially when covalent modifications are present⁴, including glycosylation^{22, 23}. However, because proteinase K is an aggressive protease it is necessary to attenuate its proteolytic activity to obtain high sequence coverage of proteins. Attenuation of PK can be achieved using suboptimal reaction conditions²¹ (to reduce the rate of enzyme activity) and limited reaction time. Using a mildly acidic, chaotrope-free solution to attenuate the activity of PK, we were able to achieve >95% sequence coverage of candidate viral spike-proteins and identify all NGS. The proteomic strategy of DeGlyPHER is conceptually similar to our previous approach³, but it is significantly faster and more sensitive.

These improvements result from three key changes to the strategy: [1] using only mass spectrometry-compatible constituents, samples are processed in a single solution except the final step of PNGase F deglycosylation; [2] reaction volumes are kept to a minimum (5–8 μ l) for increasing the rate of reaction, limit sample loss on surfaces and minimize freeze-drying time; and [3] the use of PK provides faster digestion and excellent sequence and NGS coverage with less starting material. DeGlyPHER reduces sample preparation time from 3 days to 1 day, reduces LC-MS/MS run times by 9 to 24-fold, and can achieve 90–180 times higher sensitivity than existing methods^{3, 7, 8}. In addition, we have developed the data analysis tool GlycoMSQuant which further reduces analysis times and simplifies data analysis.

We tested DeGlyPHER on BG505 SOSIP.664 MD39¹⁰, a stabilized native-like HIV Env trimer being developed for an HIV vaccination strategy²⁴ targeting germline precursors of broadly neutralizing antibodies (bNAbs) that are impacted by N-glycans. As in our previous approach³, the glycosylated peptides generated by PK digestion were sequentially deglycosylated; first with Endo H to remove high-mannose and hybrid N-glycans, and then with PNGase F, which removes all remaining N-glycans. The resulting residual masses on asparagine (N) in NGS is +203 Da at sites occupied by high-mannose/hybrid N-glycans or +3 Da at sites occupied by complex N-glycans when PNGase F deglycosylation is carried out in the presence of H₂¹⁸O (differentiating these sites from any deamidated Ns). Unoccupied NGS results in no (+0 Da) residual mass on N. Using DeGlyPHER (Figure 1a), we achieved >99% amino acid sequence coverage and identified all theoretically possible 27 NGSs from a single LC-MS/MS run of 0.5 μ g of peptides generated from a starting material of 5 μ g purified protein (Figure 1b and Figures S5a,b). We used semi-quantitative label-free analysis based on precursor peak areas to calculate the proportion of N-glycan occupancy (unoccupied: complex: high-mannose/hybrid N-glycans) for each NGS. We reanalyzed the N-glycan microheterogeneity pattern on BG505 SOSIP.664⁹ HIV Env trimer from data obtained using our previous approach³ and compared it with results using DeGlyPHER (Figure S4). The results with DeGlyPHER were highly comparable to those using our original approach, in spite of being processed differently and the samples being prepared at different times in different laboratories.

Initial results demonstrated that DeGlyPHER is at least 18 times more sensitive than our previous approach³ even though it uses a simpler and shorter workflow. To evaluate the limit of sensitivity of DeGlyPHER, we processed progressively decreasing amounts of starting material, ranging from 1 μg to 5 ng. We observed that a single LC-MS/MS run with 1 μg of starting material was enough to cover >95% of the amino acid sequence and all NGS (Figure 2a), which is 90 times more sensitive than our previous approach³. Substantial difference in microheterogeneity at any NGS, if observed, was generally when we started with <100 ng material (Figure S5c). This is likely due to low sampling as evidenced by a decrease in amino acid sequence and NGS coverage (Figure 2a), as well as the absolute number of identified peptides representing each NGS (Figure S5d).

DeGlyPHER is agnostic to mass spectrometry platform (Figure 2b). A timsTOF Pro mass-spectrometer coupled to an Evosep One HPLC (timsTOF/Evosep)²⁵ was used to achieve >99% sequence coverage and identification of all NGS using a single LC-MS/MS run with 0.5 μg of starting material and an 88-minute LC gradient (Figure 1c and Figures S6a,b). Thus, the sensitivity of DeGlyPHER on this platform was 180 times higher than our previous approach³.

N-glycan heterogeneity reflects the immunogenicity of the viral spike-protein and is critical for designing vaccines²⁶. The reproducibility of N-glycan heterogeneity patterns obtained with DeGlyPHER suggests that this is a robust procedure. Variability in sequence coverage is not observed until the limits of detection are reached on an LC-MS/MS platform. Although our results within the same LC-MS platform were reproducible (except some variations when using different proteases, Figure S3), we may infer the effects of sampling differences when comparing two different LC-MS platforms (QE-HFX/nLC vs. timsTOF/Evosep), as in case of N156, N160, N197, N386, N392 (relative peptide abundance is persistently low) and N88, N295, N301, N332, N355, N406, N411 (possible skewing of timsTOF/Evosep identification against N+203 peptides when peptide sampling per NGS decreases due to less starting material) (Figure 2b, Figures S5c,d and S6c,d). When enough sampling per NGS is achieved, these variations are diminished (Figures 1b,c). As an experiential guideline, processing 1 μg starting material in triplicates seems to be a lower limit with the current conditions we use, while still providing good sampling per NGS, for a confident data interpretation. We refrain from suggesting any sampling cutoff, as we feel that it should be determined by the users based on their data.

We attribute improvements in DeGlyPHER to efficient sample handling strategies. We observed reduced sequence coverage if the sample was from the digestion of a small amount of starting material rather than an equal aliquot from a larger sample (Figure S6e). We infer that the sensitivity differences are not occurring during LC-MS/MS, but that sample is being lost to the reaction-tube surface (during reaction and lyophilization) and the proportion of loss is more pronounced when we start with less material. The kinetics of the enzyme-substrate reaction may also account for sensitivity differences since a more “crowded” reactant environment (low reaction volumes) is expected to result in better reaction kinetics²⁷.

CONCLUSION:

The simplicity and high reproducibility of DeGlyPHER will allow for high-throughput analyses of viral spike-proteins and for any glycoprotein whether produced recombinantly or purified from natural sources. Results were highly comparable in the two LC-MS/MS platforms we used. A single analysis using a QE-HFX/nLC platform can determine the complete N-glycan heterogeneity pattern from 1 microgram of purified viral spike-protein. The timsTOF/Evosep platform was observed to be more sensitive than QE-HFX/nLC across all NGS, although limited sampling may not allow us to confidently infer N-glycan microheterogeneity at all NGS. We view the high sensitivity of DeGlyPHER to be an important step to analyze glycosylation of more complex samples such as whole virus or virus in infected blood.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS:

We thank Bruker Daltonics for providing access to timsTOF Pro MS. We thank Bruker Scientific LLC and Robin S Park for making IP2 accessible for data analysis. We thank Titus Jung in J.R.Y. lab for IT support. We thank members of the J.R.Y. lab for discussion and Claire Delahunty for critically perusing the manuscript. We thank Saman Eskandarzadeh, Michael Kubitz, and Erik Georgeson in W.R.S. lab for assistance in protein production. This work was supported by the grants P41GM103533 (NIH) and UM1AI100663, UM1AI144462, R01/R56AI113867 (NIH/NIAD). We thank Bill and Melinda Gates Foundation (BMGF) Collaboration for AIDS Vaccine Discovery (CAVD) funding to IAVI Neutralizing Antibody Center (NAC) and Ragon Institute post-doctoral fellowship to T.S.

Data Availability:

Mass spectrometry data has been deposited in MassIVE-KB repository and is also accessible through ProteomeXchange Consortium with identifiers MSV000087414 and PXD025990, respectively.

REFERENCES:

1. Rudd PM; Dwek RA, Glycosylation: heterogeneity and the 3D structure of proteins. *Crit Rev Biochem Mol Biol* 1997, 32 (1), 1–100. [PubMed: 9063619]
2. Burton DR; Hangartner L, Broadly Neutralizing Antibodies to HIV and Their Role in Vaccine Design. *Annu Rev Immunol* 2016, 34, 635–59. [PubMed: 27168247]
3. Cao L; Diedrich JK; Kulp DW; Pauthner M; He L; Park SR; Sok D; Su CY; Delahunty CM; Menis S; Andrabi R; Guenaga J; Georgeson E; Kubitz M; Adachi Y; Burton DR; Schief WR; Yates JR III; Paulson JC, Global site-specific N-glycosylation analysis of HIV envelope glycoprotein. *Nat Commun* 2017, 8, 14954. [PubMed: 28348411]
4. MacCoss MJ; McDonald WH; Saraf A; Sadygov R; Clark JM; Tasto JJ; Gould KL; Wolters D; Washburn M; Weiss A; Clark JI; Yates JR 3rd, Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci U S A* 2002, 99 (12), 7900–5. [PubMed: 12060738]
5. Zhang Y; Fonslow BR; Shan B; Baek MC; Yates JR 3rd, Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* 2013, 113 (4), 2343–94. [PubMed: 23438204]
6. Tsiatsiani L; Heck AJ, Proteomics beyond trypsin. *Febs j* 2015, 282 (14), 2612–26. [PubMed: 25823410]

7. Behrens AJ; Harvey DJ; Milne E; Cupo A; Kumar A; Zitzmann N; Struwe WB; Moore JP; Crispin M, Molecular Architecture of the Cleavage-Dependent Mannose Patch on a Soluble HIV-1 Envelope Glycoprotein Trimer. *J Virol* 2017, 91 (2).
8. Go EP; Ding H; Zhang S; Ringe RP; Nicely N; Hua D; Steinbock RT; Golabek M; Alin J; Alam SM; Cupo A; Haynes BF; Kappes JC; Moore JP; Sodroski JG; Desaire H, Glycosylation Benchmark Profile for HIV-1 Envelope Glycoprotein Production Based on Eleven Env Trimers. *J Virol* 2017, 91 (9).
9. 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, 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* 2013, 9 (9), e1003618. [PubMed: 24068931]
10. Steichen JM; Kulp DW; Tokatljan T; Escolano A; Dosenovic P; Stanfield RL; McCoy LE; Ozorowski G; Hu X; Kalyuzhnyi O; Briney B; Schiffner T; Garces F; Freund NT; Gitlin AD; Menis S; Georgeson E; Kubitz M; Adachi Y; Jones M; Mutafyan AA; Yun DS; Mayer CT; Ward AB; Burton DR; Wilson IA; Irvine DJ; Nussenzweig MC; Schief WR, HIV Vaccine Design to Target Germline Precursors of Glycan-Dependent Broadly Neutralizing Antibodies. *Immunity* 2016, 45 (3), 483–496. [PubMed: 27617678]
11. Scanlan CN; Ritchie GE; Baruah K; Crispin M; Harvey DJ; Singer BB; Lucka L; Wormald MR; Wentworth P Jr.; Zitzmann N; Rudd PM; Burton DR; Dwek RA, Inhibition of mammalian glycan biosynthesis produces non-self antigens for a broadly neutralising, HIV-1 specific antibody. *J Mol Biol* 2007, 372 (1), 16–22. [PubMed: 17631311]
12. Fan JQ; Lee YC, Detailed studies on substrate structure requirements of glycoamidases A and F. *J Biol Chem* 1997, 272 (43), 27058–64. [PubMed: 9341145]
13. He L; Diedrich J; Chu YY; Yates JR 3rd, Extracting Accurate Precursor Information for Tandem Mass Spectra by RawConverter. *Anal Chem* 2015, 87 (22), 11361–7. [PubMed: 26499134]
14. Xu T; Park SK; Venable JD; Wohlschlegel JA; Diedrich JK; Cociorva D; Lu B; Liao L; Hewel J; Han X; Wong CCL; Fonslow B; Delahunty C; Gao Y; Shah H; Yates JR 3rd, ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. *J Proteomics* 2015, 129, 16–24. [PubMed: 26171723]
15. Tabb DL; McDonald WH; Yates JR 3rd, DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* 2002, 1 (1), 21–6. [PubMed: 12643522]
16. Peng J; Elias JE; Thoreen CC; Licklider LJ; Gygi SP, Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* 2003, 2 (1), 43–50. [PubMed: 12643542]
17. Park SK; Venable JD; Xu T; Yates JR 3rd, A quantitative analysis software tool for mass spectrometry-based proteomics. *Nat Methods* 2008, 5 (4), 319–22. [PubMed: 18345006]
18. Zhang M; Gaschen B; Blay W; Foley B; Haigwood N; Kuiken C; Korber B, Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology* 2004, 14 (12), 1229–46. [PubMed: 15175256]
19. Mann HB; Whitney DR, On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics* 1947, 18 (1), 50–60.
20. Benjamini Y; Hochberg Y, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 1995, 57 (1), 289–300.
21. Wu CC; MacCoss MJ; Howell KE; Yates JR 3rd, A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* 2003, 21 (5), 532–8. [PubMed: 12692561]
22. Dalpathado DS; Irungu J; Go EP; Butnev VY; Norton K; Bousfield GR; Desaire H, Comparative glycomics of the glycoprotein follicle stimulating hormone: glycopeptide analysis of isolates from two mammalian species. *Biochemistry* 2006, 45 (28), 8665–73. [PubMed: 16834341]

23. Ruhaak LR; Xu G; Li Q; Goonatilake E; Lebrilla CB, Mass Spectrometry Approaches to Glycomic and Glycoproteomic Analyses. *Chem Rev* 2018, 118 (17), 7886–7930. [PubMed: 29553244]
24. Steichen JM; Lin YC; Havenar-Daughton C; Pecetta S; Ozorowski G; Willis JR; Toy L; Sok D; Liguori A; Kratochvil S; Torres JL; Kalyuzhnyi O; Melzi E; Kulp DW; Raemisch S; Hu X; Bernard SM; Georgeson E; Phelps N; Adachi Y; Kubitz M; Landais E; Umotoy J; Robinson A; Briney B; Wilson IA; Burton DR; Ward AB; Crotty S; Batista FD; Schief WR, A generalized HIV vaccine design strategy for priming of broadly neutralizing antibody responses. *Science* 2019, 366 (6470).
25. Meier F; Brunner AD; Koch S; Koch H; Lubeck M; Krause M; Goedecke N; Decker J; Kosinski T; Park MA; Bache N; Hoerning O; Cox J; Räther O; Mann M, Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol Cell Proteomics* 2018, 17 (12), 2534–2545. [PubMed: 30385480]
26. Seabright GE; Doores KJ; Burton DR; Crispin M, Protein and Glycan Mimicry in HIV Vaccine Design. *J Mol Biol* 2019, 431 (12), 2223–2247. [PubMed: 31028779]
27. Zhou HX; Rivas G; Minton AP, Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu Rev Biophys* 2008, 37, 375–97. [PubMed: 18573087]

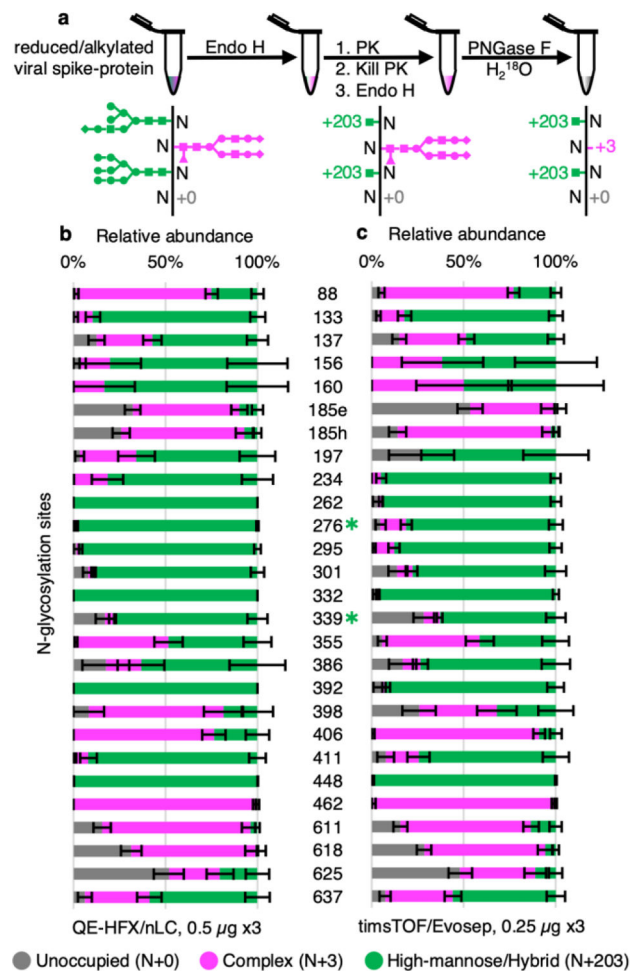


Figure 1. Proteinase K and glycosidase treatment provides N-glycan microheterogeneity in BG505 SOSIP.664 MD39 trimer. **(a)** DeGlyPHER workflow, resulting in peptides with expected mass modifications at NGS representing N-glycosylation states. **(b)** Pattern observed using 4 h triplicates on an QE-HFX/nLC platform. **(c)** Pattern observed using 88 min triplicates on timsTOF/Evosep platform. N-glycosylation states are color-coded. Error bars represent mean \pm SEM. Between **(b)** and **(c)**, the pattern is similar, and any significant difference (BH-corrected p -value <0.05) in proportion of a certain N-glycosylation state at an NGS is represented by color-coded *.

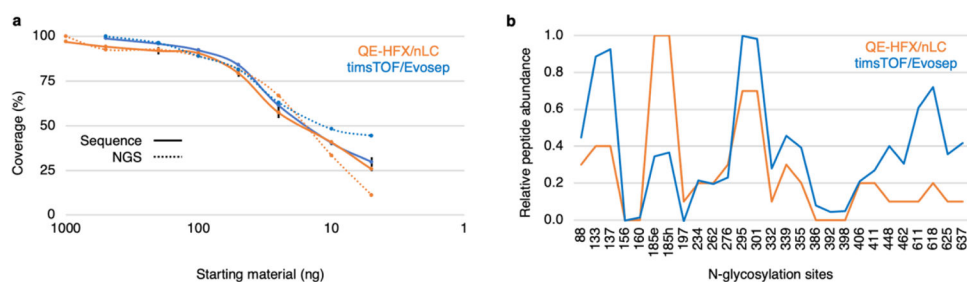


Figure 2. Factors affecting limit of sensitivity. **(a)** Decreasing trend of sequence and NGS coverage observed as starting material is diluted 200-fold (1 μ g to 5 ng, QE-HFX/nLC) or 100-fold (0.5 μ g to 5 ng, timsTOF/Evosep), using triplicates for each amount of starting material. The limit of sensitivity is revealed. **(b)** Relative abundance of peptides identified per NGS across the dilution series for the 2 LC-MS/MS platforms used. This comparison reveals a non-uniform digestion pattern that may be attributed to steric hindrance offered by the glycoprotein or characteristic behavior of individual peptides in LC-MS/MS. Error bars represent mean \pm SEM. Values for sequence and NGS coverage are mean and union across replicates, respectively.