Title:

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

Authors:

Sabyasachi Baboo*, Jolene K Diedrich, Salvador Martínez-Bartolomé, Xiaoning Wang, Torben Schiffner, Bettina Groschel, William R Schief, James C Paulson & John R Yates III*

Affiliations:

Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA, USA

Sabyasachi Baboo, Jolene K Diedrich, Salvador Martínez-Bartolomé, Xiaoning Wang, James C Paulson and John R Yates III

Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA, USA

Torben Schiffner, Bettina Groschel, William R Schief, James C Paulson

IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, CA, USA

Torben Schiffner, Bettina Groschel, William R Schief

The Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard, Cambridge, MA, USA

Torben Schiffner, William R Schief

Present address:

Institut für Wirkstoffentwicklung, Universität Leipzig, Leipzig, Germany

Torben Schiffner

Corresponding authors:

Correspondence to Sabyasachi Baboo (sbaboo@scripps.edu) or John R Yates III (jyates@scripp.sedu)

Keywords:

proteomics, mass spectrometry, Proteinase K, N-glycan, heterogeneity, HIV, *Env*, viral spike

Abstract:

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.

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 response[1]. These glycans add complexity to the production and characterization of recombinant protein-based vaccines[2]. 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

47 the sequon NX[S|T], where X is any amino acid except P[3]. To address this analytical challenge
48 several mass spectrometry-based strategies using multiple proteases[4-6] have been implemented to
49 create sufficient numbers of peptides unique to each glycosylation site[3,7,8]. In these strategies,
50 individual aliquots are digested with each protease and analyzed separately by liquid
51 chromatography-mass spectrometry (LC-MS/MS), or pooled and analyzed together. To broadly
52 characterize the nature of the glycosylation at each NGS, we had previously introduced the use
53 of endoglycosidases that create residual mass signatures[3]. This helped us to determine the degree
54 of glycan occupancy, and the degree of glycan processing – from the high mannose form that is
55 initially attached to the protein, which may mature into the complex form when mannose
56 residues are replaced by "terminal" monosaccharide sequences. To achieve coverage for all
57 NGS, we had combined several digestions performed with different proteases and achieved
58 >95% sequence coverage. Here we show that it is possible to replace these multiple proteolytic
59 digestions with a single Proteinase K (PK) digestion, and moreover through careful choice of
60 volatile buffers we have developed an improved "single pot" strategy with significantly
61 increased sensitivity. We name this strategy to analyze glycoforms as DeGlyPHER
62 (*De*glycosylation-dependent *Gly*can/*P*roteomic *H*eterogeneity *E*valuation *R*eport).
63

64 **Results and Discussion:**

65 PK is a broadly specific serine-protease that has previously been exploited for its potential to
66 generate overlapping peptides and high sequence coverage[9]. The redundancy afforded by
67 overlapping sequences significantly increases confidence in identifications, especially when
68 covalent modifications are present[4]. However, because proteinase K is an aggressive protease it
69 is necessary to attenuate its proteolytic activity to obtain high sequence coverage of proteins.

70 Attenuation of PK can be achieved using suboptimal reaction conditions[9] (to reduce the rate of
71 enzyme activity) and limited reaction time. Using a mildly acidic, chaotrope-free solution to
72 attenuate the activity of PK, we were able to achieve >95% sequence coverage of candidate viral
73 spike-proteins and identify all NGS. The proteomic strategy of DeGlyPHER is conceptually
74 similar to our previous approach[3], but it is significantly faster and more sensitive. These
75 improvements result from three key changes to the strategy: [1] using only mass spectrometry-
76 compatible constituents, samples are processed in a single solution except the final step of
77 PNGase F deglycosylation; [2] reaction volumes are kept to a minimum (5-8 µl) to increase the
78 rate of reaction to limit sample loss on surfaces and minimize freeze-drying time; and [3] the use
79 of PK provides faster digestion and excellent sequence and NGS coverage with less starting
80 material. DeGlyPHER reduces sample preparation time from 3 days to 1 day, reduces LC-
81 MS/MS run times by 9 to 24-fold, and can achieve 90-180 times higher sensitivity than existing
82 methods[3,7,8]. In addition, we have developed the data analysis tool GlycoMSQuant which further
83 reduces analysis times and simplifies data analysis.
84
85 We tested DeGlyPHER on BG505 SOSIP.664 MD39[10], a stabilized native-like HIV *Env* trimer
86 being developed for an HIV vaccination strategy[11] targeting germline precursors of broadly
87 neutralizing antibodies (bNAbs) that are impacted by N-glycans. As in our previous approach[3],
88 the glycosylated peptides generated by PK digestion were sequentially deglycosylated; first with
89 Endo H to remove high-mannose and hybrid N-glycans, and then with PNGase F, which
90 removes all remaining N-glycans. The resulting residual masses on asparagine (N) in NGS is
91 +203 Da at sites occupied by high-mannose/hybrid N-glycans or +3 Da at sites occupied by
92 complex N-glycans when PNGase F deglycosylation is carried out in the presence of $H_2{}^{18}O$

93 (differentiating these sites from any deamidated Ns). Unoccupied NGS results in no (+0 Da)
94 residual mass on N. Using DeGlyPHER (***Fig. 1a***), we achieved >99% amino acid sequence
95 coverage and identified all theoretically possible 27 NGSs from a single LC-MS/MS run of 0.5
96 µg of peptides generated from a starting material of 5 µg purified protein (***Fig. 1b*** and
97 ***Supplementary Figures 1a,b***). We used semi-quantitative label-free analysis based on precursor
98 peak areas to calculate the proportion of N-glycan occupancy (unoccupied: complex: high-
99 mannose/hybrid N-glycans) for each NGS. We reanalyzed the N-glycan microheterogeneity
100 pattern on BG505 SOSIP.664[12] HIV *Env* trimer from data obtained using our previous approach[3]
101 and compared it with results using DeGlyPHER (***Supplementary Figures 2***). The results with
102 DeGlyPHER were highly comparable to those using our original approach, in spite of being
103 processed differently and the samples being prepared at different times in different laboratories.
104
105 Initial results demonstrated that DeGlyPHER is at least 18 times more sensitive than our
106 previous approach[3] even though it uses a simpler and shorter workflow. To evaluate the limit of
107 sensitivity of DeGlyPHER, we processed progressively decreasing amounts of starting material,
108 ranging from 1 µg to 5 ng. We observed that a single LC-MS/MS run with 1 µg of starting
109 material was enough to cover >95% of the amino acid sequence and all NGS (***Fig. 2a***), which is
110 90 times more sensitive than our previous approach[3]. Major differences in microheterogeneity at
111 each NGS were generally observed when we started with <100 ng material (***Supplementary***
112 ***Figure 1c***). This is likely due to low sampling as evidenced by a decrease in amino acid
113 sequence and NGS coverage (***Fig. 2a***), as well as the absolute number of identified peptides
114 representing each NGS (***Supplementary Figure 1d***).
115

116 DeGlyPHER is agnostic to mass spectrometry platform (***Fig. 2b***). A timsTOF Pro mass-
117 spectrometer coupled to an Evosep One HPLC (timsTOF/Evosep)[13] was used to achieve >99%
118 sequence coverage and identification of all NGS using a single LC-MS/MS run with 0.5 µg of
119 starting material and an 88-minute LC gradient (***Fig. 1c*** and ***Supplementary Figures 3a,b***).
120 Thus, the sensitivity of DeGlyPHER on this platform was 180 times higher than our previous
121 approach[3].
122
123 N-glycan heterogeneity reflects the immunogenicity of the viral spike-protein and is critical for
124 designing vaccines[14]. The reproducibility of N-glycan heterogeneity patterns obtained with
125 DeGlyPHER suggests that this is a robust procedure. Variability in sequence coverage is not
126 observed until the limits of detection are reached on an LC-MS/MS platform. Although our
127 results within the same LC-MS platform were reproducible (except some variations when using
128 different proteases, ***Supplementary Figure 5***), we may infer the effects of sampling differences
129 when comparing two different LC-MS platforms (QE-HFX *vs.* timsTOF/Evosep), as in case of
130 N156, N160, N197, N386, N392 (relative peptide abundance is persistently low) and N88, N295,
131 N301, N332, N355, N406, N411 (possible skewing of timsTOF/Evosep identification against
132 N+203 peptides when peptide sampling per NGS decreases due to less starting material) (***Fig. 2b***
133 and ***Supplementary Figures 1c,d*** and ***3c,d***). When enough sampling per NGS is achieved, these
134 variations are diminished (***Figs. 1b,c***).
135
136 We attribute improvements in DeGlyPHER to efficient sample handling strategies. We observed
137 reduced sequence coverage if the sample was from the digestion of a small amount of starting
138 material rather than an equal aliquot from a larger sample (***Supplementary Figure 3e***). We infer

139 that the sensitivity differences are not occurring during LC-MS/MS, but that sample is being lost
140 to the reaction-tube surface (during reaction and lyophilization) and the proportion of loss is
141 more pronounced when we start with less material. The kinetics of the enzyme-substrate reaction
142 may also account for sensitivity differences since a more "crowded" reactant environment (low
143 reaction volumes) is expected to result in better reaction kinetics[15].
144
145 The simplicity and high reproducibility of DeGlyPHER will allow for high-throughput analyses
146 of viral spike-proteins and for any glycoprotein whether produced recombinantly or purified
147 from natural sources. Results were highly comparable in the two LC-MS/MS platforms we used.
148 A single analysis using a QE-HFX/nLC platform can determine the complete N-glycan
149 heterogeneity pattern from 1 microgram of purified viral spike-protein. The timsTOF/Evosep
150 platform was observed to be more sensitive than QE-HFX/nLC across all NGS, although limited
151 sampling may not allow us to confidently infer N-glycan microheterogeneity at all NGS. We
152 view the high sensitivity of DeGlyPHER to be an important step to analyze glycosylation of
153 more complex samples such as whole virus or virus in infected blood[16].
154
155 **Methods:**
156 Expression and purification of HIV *Env* trimers
157 BG505 SOSIP.664[12] and BG505 SOSIP.664 MD39[10] *Env* trimers were expressed and purified
158 essentially as described previously[10]. Briefly, sequences with codons optimized for expression in
159 human cells were synthesized and cloned into pHLSec between *AgeI*/*KpnI* by Genscript. The
160 constructs were co-transfected with Furin-encoding plasmid, using polyethylenimine in Freestyle
161 293F cells cultured in 293 FreeStyle media (Thermo Fisher Scientific). Where indicated, 15 μM

162 sterile-filtered Kifunensine (Cayman Chemical) was added after transfection. After 6-7 days,
163 supernatant was collected after passing through 0.22 µm filter (Nalgene), and the C-terminally
164 His-tagged trimers were purified using a HisTrap affinity column (Cytiva) with a linear elution
165 gradient from 20-500 mM imidazole, followed by a Superdex 200 Increase SEC column (Cytiva)
166 in Tris-buffered saline/TBS (20 mM Tris, 100 mM NaCl, pH 7.5). The oligomeric state and
167 purity of trimer was verified using size exclusion chromatography coupled with multi-angle light
168 scattering (SEC-MALS; DAWN HELEOS II/ Optilab T-rEX, Wyatt Technology).
169
170 Proteinase K treatment and deglycosylation
171 HIV Env trimer was exchanged to water using Microcon Ultracel-10 centrifugal device
172 (Millipore Sigma). Trimer was reduced with 5 mM tris(2-carboxyethyl)phosphine hydrochloride
173 (TCEP-HCl, Thermo Scientific) and alkylated with 10 mM 2-chloroacetamide (Sigma Aldrich)
174 in 100 mM ammonium acetate for 20 min at room temperature (RT, 24°C). Initial protein-level
175 deglycosylation was performed using 250 U of Endo H (New England Biolabs) for up to 5 µg
176 trimer, for 1 h at 37°C (pH 5.5-6.0). Trimer was digested with 1:25 Proteinase K (Sigma
177 Aldrich) for 4 h at 37°C (pH 5.5-6.0). PK was denatured by incubating at 90°C for 15 min, then
178 cooled to RT. Peptides were deglycosylated again with 250 U Endo H for 1 h at 37°C (pH 5.5-
179 6.0), then frozen at –80°C and lyophilized. 100 U PNGase F (New England Biolabs) was
180 lyophilized (for up to 5 µg trimer), resuspended in 100 mM ammonium bicarbonate prepared in
181 $H_2{}^{18}O$ (97% ^{18}O, Sigma-Aldrich), and added to the lyophilized peptides. The resulting 5-8 µl
182 reaction solutions (except PNGase F reaction, in 5-20 µl) were then incubated for 1 h at 37°C
183 (pH 8.0-8.5) in 0.2 ml PCR tubes on a thermocycler with heated lid.
184

185 Validating efficiency of glycosidases

186 BG505 SOSIP.664 HIV *Env* trimer glycosylated with only high-mannose N-glycans (purified
187 from cells treated with Kifunensine[17], which inhibits processing of high-mannose N-glycans to
188 complex N-glycans during protein maturation), was sequentially treated with Endo H, followed
189 by PNGase F. After treatment with both enzymes, 99.2% of identified peptides were N+203;
190 100% of peptides identified with only PNGase F treatment were N+3 (***Supplementary Figures***
191 ***4a,b***; proportions do not consider unoccupied NGS because they remained similar in both
192 experiments – 8-10%). We realize the possibility that glycosidase PNGase F may occasionally
193 cleave the remnant GlcNAc (N-Acetylglucosamine)[18] post-Endo H processing of high-
194 mannose/hybrid N-glycans and thus, convert the mass modification characteristic of high-
195 mannose/hybrid (+203) to complex (+3) N-glycans, which would affect our analyses. However,
196 we have not observed any significant evidence of this possibility in our results, though it may
197 explain why we observe a few peptides with +3 mass modified NGS in Kifunensine treated
198 samples (***Supplementary Figure 4a***).

199

200 Trypsin proteolysis

201 The Proteinase K/deglycosylation method described above was followed, except PK was
202 replaced with trypsin and reactions were incubated overnight at 37°C. Trypsin generated a lower
203 total number of peptides than PK, but we obtained >95% sequence coverage, including 26 of 27
204 NGS (***Supplementary Figure 5a***). Variations in N-glycan microheterogeneity at certain NGS
205 may be explained by low sampling at these sites (N386, N392) or difference in cleavage-
206 specificity between PK and trypsin (N88, N611) (***Supplementary Figures 5b,c***).

207

208 LC-MS/MS

209 *Q Exactive HF-X with EASY-nLC 1200*

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

222 *timsTOF Pro with Evosep One*

223 Samples were loaded onto EvoTips following manufacturer protocol. The samples were run on
224 an Evosep One (Evosep) coupled to a timsTOF Pro (Bruker Daltonics). Samples were separated
225 on a 15 cm × 150 µm ID column with BEH 1.7 µm C18 beads (Waters) and integrated tip pulled
226 in-house using either the 30 SPD or 15 SPD methods. Mobile phases A and B were 0.1% formic
227 acid in water and 0.1% formic acid in acetonitrile, respectively. MS data was acquired in PASEF
228 mode with 1 MS1 survey TIMS-MS and 10 PASEF MS/MS scans acquired per 1.1 s acquisition
229 cycle. Ion accumulation and ramp time in the dual TIMS analyzer was set to 100 ms each and we
230 analyzed the ion mobility range from $1/K_0 = 0.6$ Vs cm^{-2} to 1.6 Vs cm^{-2}. Precursor ions for

231 MS/MS analysis were isolated with a 2 Th window for m/z < 700 and 3 Th for m/z >700 with a
232 total m/z range of 100-1700. The collision energy was lowered linearly as a function of
233 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}.
234 Singly charged precursor ions were excluded with a polygon filter, precursors for MS/MS were
235 picked at an intensity threshold of 2,500, target value of 20,000 and with an active exclusion of
236 24 s.
237
238 Data Processing
239 Protein and peptide identification were done with Integrated Proteomics Pipeline (IP2, Bruker
240 Scientific LLC). Tandem mass spectra were extracted from raw files using RawConverter[19]
241 (timstofCoverter for timsTOF Pro data) and searched with ProLuCID[20] against a database
242 comprising UniProt reviewed (Swiss-Prot) proteome for *Homo sapiens* (UP000005640), UniProt
243 amino acid sequences for Endo H (P04067), PNGase F (Q9XBM8), and Proteinase K (P06873),
244 amino acid sequences for BG505 SOSIP.664[12] and BG505 SOSIP.664 MD39[10] (including a
245 preceding secretory signal sequence and followed by 6xHis-tag), and a list of general protein
246 contaminants. The search space included no cleavage-specificity (all fully tryptic and semi-
247 tryptic peptide candidates for trypsin treatment). Carbamidomethylation (+57.02146 C) was
248 considered a static modification. Deamidation in presence of $H_2{}^{18}O$ (+2.988261 N), GlcNAc
249 (+203.079373 N), oxidation (+15.994915 M) and N-terminal pyroglutamate formation (–
250 17.026549 Q) were considered differential modifications. Data was searched with 50 ppm
251 precursor ion tolerance and 50 ppm fragment ion tolerance. Identified proteins were filtered
252 using DTASelect2[21] and utilizing a target-decoy database search strategy to limit the false
253 discovery rate to 1%, at the spectrum level[22]. 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). Census2[23] 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 GlycoMSQuant 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 HXB2[24] HIV-1 variant.

Precursor peak area was calculated by Census2[23] 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 ∈ *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 test[25]. Testing $\%_{g,pepz,a}$ *vs.* $\%_{g,pepz,b}$ individually for *u*, *c* and *h* at each NGS, we

298 calculated *p*-values that were subjected to multiple hypothesis correction using the Benjamini-
299 Hochberg (BH) method[26]. If the corrected *p*-value was <0.05 (for *u*, *c* or *h* at any NGS), then the
300 difference was considered statistically significant.
301
302 For ***Supplementary Figure 2***, published data from our previous approach[3] was reanalyzed using
303 the data analysis workflow described here. Briefly, the data analyzed is from 3 replicates of 3
304 conditions, each separately analyzed by LC-MS/MS, with total starting material of 90 µg protein,
305 and Census2[23] label-free analysis was performed simultaneously on all 9 experiments without
306 "match between runs", and the results analyzed by GlycoMSQuant. This was compared with
307 data obtained from a single LC-MS/MS run (QE-HFX/nLC) with 0.5 µg of peptides generated
308 from a starting material of 5 µg purified protein.
309
310 **Acknowledgement:**
311 We thank Bruker Daltonics for providing access to timsTOF Pro MS. We thank Bruker
312 Scientific LLC and Robin S Park for making IP2 accessible for data analysis. We thank Titus
313 Jung in J.R.Y. lab for IT support. We thank members of the J.R.Y. lab for discussion and Claire
314 Delahunty for critically perusing the manuscript. We thank Saman Eskandarzadeh, Michael
315 Kubitz, and Erik Georgeson in W.R.S. lab for assistance in protein production. This work was
316 supported by the grants P41GM103533 (NIH) and UM1AI100663, UM1AI144462,
317 R01/R56AI113867 (NIH/NIAID). We thank Bill and Melinda Gates Foundation (BMGF)
318 Collaboration for AIDS Vaccine Discovery (CAVD) funding to IAVI Neutralizing Antibody
319 Center (NAC) and Ragon Institute post-doctoral fellowship to T.S.
320

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.

Competing Interests:

The authors declare no competing interests.

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.

Code Availability:

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

References:

1 Rudd, P. M. & Dwek, R. A. Glycosylation: heterogeneity and the 3D structure of proteins. *Crit Rev Biochem Mol Biol* **32**, 1-100, doi:10.3109/10409239709085144 (1997).

342 2 Burton, D. R. & Hangartner, L. Broadly Neutralizing Antibodies to HIV and Their Role
343 in Vaccine Design. *Annu Rev Immunol* **34**, 635-659, doi:10.1146/annurev-immunol-
344 041015-055515 (2016).
345 3 Cao, L. *et al.* Global site-specific N-glycosylation analysis of HIV envelope glycoprotein.
346 *Nat Commun* **8**, 14954, doi:10.1038/ncomms14954 (2017).
347 4 MacCoss, M. J. *et al.* Shotgun identification of protein modifications from protein
348 complexes and lens tissue. *Proc Natl Acad Sci U S A* **99**, 7900-7905,
349 doi:10.1073/pnas.122231399 (2002).
350 5 Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C. & Yates, J. R., 3rd. Protein analysis by
351 shotgun/bottom-up proteomics. *Chem Rev* **113**, 2343-2394, doi:10.1021/cr3003533
352 (2013).
353 6 Tsiatsiani, L. & Heck, A. J. Proteomics beyond trypsin. *Febs j* **282**, 2612-2626,
354 doi:10.1111/febs.13287 (2015).
355 7 Behrens, A. J. *et al.* Molecular Architecture of the Cleavage-Dependent Mannose Patch
356 on a Soluble HIV-1 Envelope Glycoprotein Trimer. *J Virol* **91**, doi:10.1128/jvi.01894-16
357 (2017).
358 8 Go, E. P. *et al.* Glycosylation Benchmark Profile for HIV-1 Envelope Glycoprotein
359 Production Based on Eleven Env Trimers. *J Virol* **91**, doi:10.1128/jvi.02428-16 (2017).
360 9 Wu, C. C., MacCoss, M. J., Howell, K. E. & Yates, J. R., 3rd. A method for the
361 comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* **21**, 532-538,
362 doi:10.1038/nbt819 (2003).

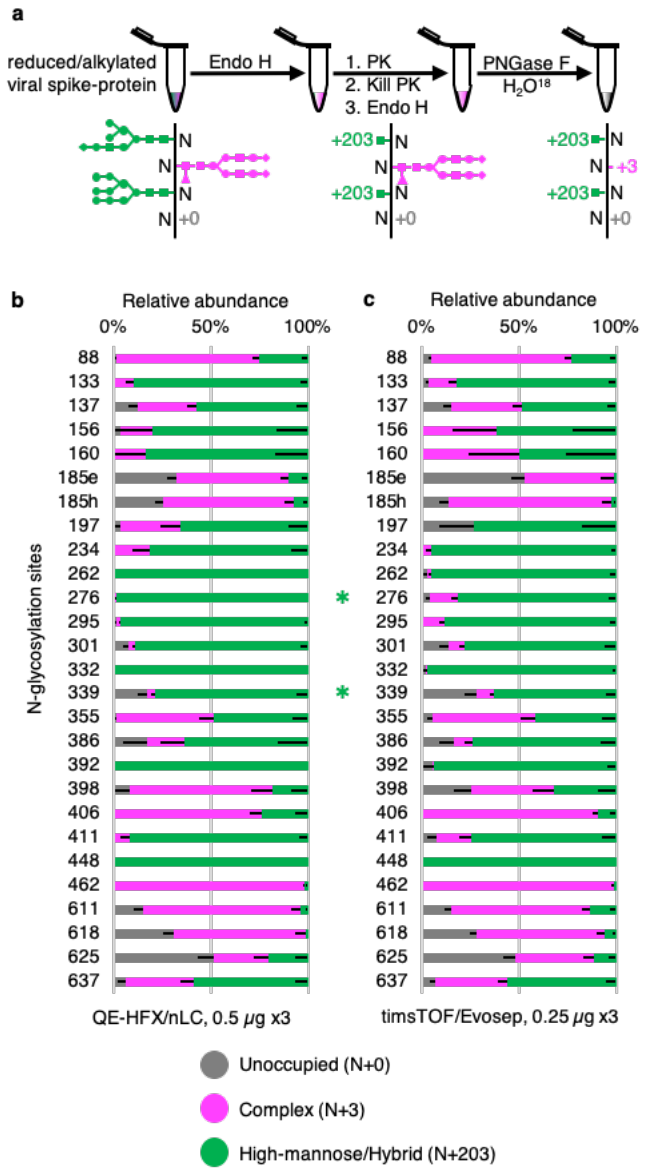
363 10 Steichen, J. M. *et al.* HIV Vaccine Design to Target Germline Precursors of Glycan-
364 Dependent Broadly Neutralizing Antibodies. *Immunity* **45**, 483-496,
365 doi:10.1016/j.immuni.2016.08.016 (2016).
366 11 Steichen, J. M. *et al.* A generalized HIV vaccine design strategy for priming of broadly
367 neutralizing antibody responses. *Science* **366**, doi:10.1126/science.aax4380 (2019).
368 12 Sanders, R. W. *et al.* A next-generation cleaved, soluble HIV-1 Env trimer, BG505
369 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-
370 neutralizing antibodies. *PLoS Pathog* **9**, e1003618, doi:10.1371/journal.ppat.1003618
371 (2013).
372 13 Meier, F. *et al.* Online Parallel Accumulation-Serial Fragmentation (PASEF) with a
373 Novel Trapped Ion Mobility Mass Spectrometer. *Mol Cell Proteomics* **17**, 2534-2545,
374 doi:10.1074/mcp.TIR118.000900 (2018).
375 14 Seabright, G. E., Doores, K. J., Burton, D. R. & Crispin, M. Protein and Glycan Mimicry
376 in HIV Vaccine Design. *J Mol Biol* **431**, 2223-2247, doi:10.1016/j.jmb.2019.04.016
377 (2019).
378 15 Zhou, H. X., Rivas, G. & Minton, A. P. Macromolecular crowding and confinement:
379 biochemical, biophysical, and potential physiological consequences. *Annu Rev Biophys*
380 **37**, 375-397, doi:10.1146/annurev.biophys.37.032807.125817 (2008).
381 16 Bar-On, Y. M., Flamholz, A., Phillips, R. & Milo, R. SARS-CoV-2 (COVID-19) by the
382 numbers. *Elife* **9**, doi:10.7554/eLife.57309 (2020).
383 17 Scanlan, C. N. *et al.* Inhibition of mammalian glycan biosynthesis produces non-self
384 antigens for a broadly neutralising, HIV-1 specific antibody. *J Mol Biol* **372**, 16-22,
385 doi:10.1016/j.jmb.2007.06.027 (2007).

386 18 Fan, J. Q. & Lee, Y. C. Detailed studies on substrate structure requirements of
387 glycoamidases A and F. *J Biol Chem* **272**, 27058-27064, doi:10.1074/jbc.272.43.27058
388 (1997).
389 19 He, L., Diedrich, J., Chu, Y. Y. & Yates, J. R., 3rd. Extracting Accurate Precursor
390 Information for Tandem Mass Spectra by RawConverter. *Anal Chem* **87**, 11361-11367,
391 doi:10.1021/acs.analchem.5b02721 (2015).
392 20 Xu, T. *et al.* ProLuCID: An improved SEQUEST-like algorithm with enhanced
393 sensitivity and specificity. *J Proteomics* **129**, 16-24, doi:10.1016/j.jprot.2015.07.001
394 (2015).
395 21 Tabb, D. L., McDonald, W. H. & Yates, J. R., 3rd. DTASelect and Contrast: tools for
396 assembling and comparing protein identifications from shotgun proteomics. *J Proteome*
397 *Res* **1**, 21-26, doi:10.1021/pr015504q (2002).
398 22 Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J. & Gygi, S. P. Evaluation of
399 multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-
400 MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* **2**, 43-50,
401 doi:10.1021/pr025556v (2003).
402 23 Park, S. K., Venable, J. D., Xu, T. & Yates, J. R., 3rd. A quantitative analysis software
403 tool for mass spectrometry-based proteomics. *Nat Methods* **5**, 319-322,
404 doi:10.1038/nmeth.1195 (2008).
405 24 Zhang, M. *et al.* Tracking global patterns of N-linked glycosylation site variation in
406 highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza
407 hemagglutinin. *Glycobiology* **14**, 1229-1246, doi:10.1093/glycob/cwh106 (2004).

25 Mann, H. B. & Whitney, D. R. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics* **18**, 50-60 (1947).

26 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)* **57**, 289-300 (1995).

431 **Figures**



432

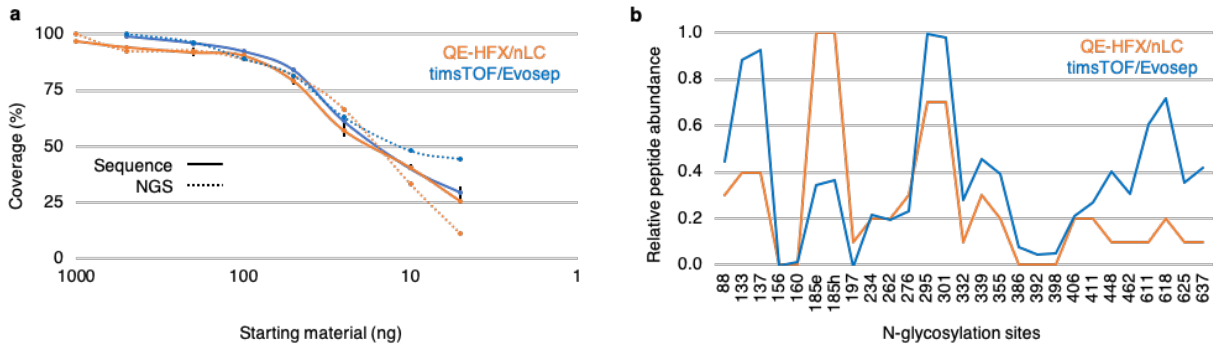
433 **Figure 1 | Proteinase K and glycosidase treatment provides N-glycan microheterogeneity in BG505 SOSIP.664 MD39 trimer.** (**a**) DeGlyPHER workflow, resulting in peptides with expected mass modifications at NGS

434 representing N-glycosylation states. (**b**) Pattern observed using 4 h triplicates on an QE-HFX/nLC platform. (**c**) Pattern observed using 88 min

435 triplicates on timsTOF/Evosep platform. N-glycosylation states are color-coded. Error bars represent mean-SEM. Between (**b**) and (**c**), the pattern is similar, and any significant difference (BH-corrected *p*-value <0.05) in

436 proportion of a certain N-glycosylation state at an NGS is represented by color-coded ∗.

437



438

439

440

Figure 2 I Factors affecting limit of sensitivity. (**a**) Decreasing trend of sequence and NGS coverage observed as starting material is diluted 200-fold (1 μg to 5 ng, QE-HFX/nLC) or 100-fold (0.5 μg 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±SEM. Values for sequence and NGS coverage are mean